AN INTRODUCTION
TO THE
THEORY OF OPTICS

BY

SIR ARTHUR SCHUSTER,

PH.D. (HEIDELBERG), SC.D. (CANTAB.), D.SC. (MANCHESTER, OXON. AND CALCUTTA), LL.D. (ST ANDREWS), D.-ÈS-SC. (GENEVA), F.R.S.

HON. PROFESSOR OF PHYSICS AT THE UNIVERSITY OF MANCHESTER.

THIRD EDITION

REVISED AND ENLARGED BY THE AUTHOR AND

JOHN WILLIAM NICHOLSON,

M.A. (OXON., CAMB.), D.SC. (LOND.), M.SC. (MANCH.), F.R.S.

FELLOW AND TUTOR OF BALLIOL COLLEGE, OXFORD

SOMETIME PROFESSOR OF MATHEMATICS IN THE UNIVERSITY OF LONDON

LONDON

EDWARD ARNOLD & CO.

1924

THIS VOLUME
IS DEDICATED TO

JOHN WILLIAM STRUTT

THIRD BARON RAYLEIGH

O.M., Sc.D., F.R.S.

WHO BY HIS WRITINGS HAS ADDED CLEARNESS AND PRECISION
TO NEARLY ALL BRANCHES OF OPTICS

PREFACE TO THIRD EDITION

DURING the fourteen years that have passed since the appearance of the second edition of this treatise, the quantum theory has found much support from those parts of Optics which are intimately connected with the structure of atoms. It is hoped that the two Chapters which we have consequently added may serve as an introduction to the study of this new and important branch of Physics.

The book has been thoroughly revised and additions have been made, more especially with regard to the various applications of interference phenomena. Professor Michelson's method of measuring the diameter of the stars has a great future before it, and his device for obtaining accurate values of the tidal distortion of the earth also deserved a place in the new matter which has been added to the volume.

ARTHUR SCHUSTER.
J. W. NICHOLSON.

November, 1923.

PREFACE TO SECOND EDITION.

THE changes introduced into this edition are mainly confined to correcting minor errors and removing obscurities of expression; a few more serious alterations and some additions, however, have also been made. The most important of these concerns the treatment of molecular scattering, which leads to a formula connecting the coefficient of extinction with the refractive index. This formula, which was first given by Lord Rayleigh, appeared already in the first edition, but its importance and great generality was not sufficiently emphasized.

In view of the recent progress made in establishing accurate standards of wave-length, and connecting the metre with the length of a homogeneous wave, it seemed advisable to explain the methods of measurement in some detail, and to give greater prominence to the work of Fabry and Perot.

Finally the treatment of white light and of interference problems has been made more consistent—and I hope clearer—by introducing the theory of impulses at an earlier stage. As there is still a certain hesitation to recognize that white light can be treated as an entity, and that to decompose it into homogeneous components may obscure the problem, I have tried to remove one of the stumbling-blocks by defining the word "interference" in a manner which allows a definite meaning to be attached to the expression. It is curious that in hardly any treatise on Optics is there any proper definition of the word. Mascart contents himself with saying that the principle of interference is synonymous with the principle of superposition; but if we adopt this view, we must admit that rays of light coming from two different sources interfere with each other, which is contrary to the statement universally met with in our text-books. It is better to retain in accordance with usage a distinction of meaning between two words which are both useful, and I hope that the definition I have given may prove generally acceptable. To be consistent I have had to change the title of Chapter IV, which in the previous edition carried the heading "Interference of Light." I have also ventured to introduce a new term, "quasi-homogeneous" light, to describe such approach to homogeneous light as is at our command, restricting the word "homogeneous" to the ideal case of an indefinitely extended perfectly regular oscillation. Homogeneous light is necessarily polarized, and two homogeneous rays produce interference effects though they come from different sources. If our experiments show that this is apparently not the case, it proves that the light we experiment with is not homogeneous. Yet it is well to have a word which allows us to characterize the radiations of a luminous gas, which often show great regularity, and the term I propose seems appropriate.

I have resisted the temptation to add to the later Chapters, which I know are incomplete, especially in the treatment of the effects of motion; but—as the title indicates—the book is intended for an introduction to a theory, the later developments of which are best studied by consulting the original sources.

My thanks are due to various friends and correspondents who have kindly pointed out a number of errors, which were left standing in the previous edition—but I feel a consoling though unmerited sense of satisfaction at the one serious blunder having remained unnoticed and, I hope, undetected.

ARTHUR SCHUSTER.

August, 1909.

PREFACE TO FIRST EDITION.

THERE is at present no theory of Optics in the sense that the elastic solid theory was accepted fifty years ago. We have abandoned that theory, and learned that the undulations of light are electromagnetic waves differing only in linear dimensions from the disturbances which are generated by oscillating electric currents or moving magnets. But so long as the character of the displacements which constitute the waves remains undefined we cannot pretend to have established a theory of light. This limitation of our knowledge, which in one sense is a retrogression from the philosophic standpoint of the founders of the undulatory theory, is not always sufficiently recognized and sometimes deliberately ignored. Those who believe in the possibility of a mechanical conception of the universe and are not willing to abandon the methods which from the time of Galileo and Newton have uniformly and exclusively led to success, must look with the gravest concern on a growing school of scientific thought which rests content with equations correctly representing numerical relationships between different phenomena, even though no precise meaning can be attached to the symbols used. The fact that this evasive school of philosophy has received some countenance from the writings of Heinrich Hertz renders it all the more necessary that it should be treated seriously and resisted strenuously.

The equations which at present represent the electromagnetic theory of light have rendered excellent service, and we must look upon them as a framework into which a more complete theory must necessarily fit, but they cannot be accepted as constituting in themselves a final theory of light.

The study of Physics must be based on a knowledge of Mechanics, and the problem of light will only be solved when we have discovered the mechanical properties of the æther. While we are in ignorance on fundamental matters concerning the origin of electric and magnetic strains and stresses, it is necessary to introduce the theoretical study of light by a careful treatment of wave propagation through media the elastic properties of which are known. A study of the theory of sound and of the old elastic solid theory of light must precede therefore the introduction of the electromagnetic equations.

The present volume is divided into two parts; the first part includes those portions of the subject which may be treated without the help of the equations of dynamics, although a short discussion of the kinetics of wave motion is introduced at an early stage. The mathematical treatment has been kept as simple as possible, elementary methods only being used. I hope that rigidity of method is nowhere sacrificed thereby, while the advantage is gained that students obtain an insight into what is most essential in the theory of Interference and Diffraction, without introducing purely mathematical difficulties such as are involved in the use of Fresnel's integrals. Even accurate numerical results may be obtained by a proper use of Fresnel's zones.

The second part of the book is intended to serve as an introduction to the higher branches of the subject. It has not been my object as regards this more advanced portion to write a treatise which shall be complete in itself, but rather to introduce the student to the writings of the original authorities. As a teacher, I consider this to be the correct method, being convinced that students should be encouraged at an early stage to consult the literature of the subject. It is a necessary consequence of the point of view adopted that the treatment is somewhat unequal. Where the author has nothing to say which is novel, or may remove obscurities, the best thing he can do, is to content himself with a short summary, referring the reader for details to the available sources of information. A more lengthy exposition is justified where a simplification or some new matter can be introduced. It may be mentioned in this connexion that as far as I know the consideration of absorptive regions of finite range of frequency in the theory of selective dispersion is new, and has not previously been published.

I have purposely abstained from entering into details of methods of observation or instrumental appliances. These belong more properly to the courses of laboratory instruction.

I hope that the short biographical notices of deceased authors who have made important contributions to the science will be found to be of interest.

The greater part of this book was already in type when Lord Kelvin's *Baltimore Lectures* appeared; I was still able to add some references to these lectures, though not to the extent I should have wished. In some of the later chapters repeated reference is made to Drude's *Lehrbuch der Optik.* Students who desire to pursue the subject further, should also have access to Mascart's *Optique* and Lord Rayleigh's *Collected Works.* My own indebtedness to Lord Rayleigh's writings and personal inspiration is greater than can be acknowledged by mere references to his papers, and I am therefore glad to be allowed to dedicate this volume to him.

I am obliged to Prof. Wilberforce and Mr W. H. Jackson for having looked through the proofs of the greater portion of the work, and favoured me with their corrections and suggestions. I have also to thank Mr J. E. Petavel for the very valuable help he gave me in drawing out the figures, and Mr H. E. Wood for taking the photographs of interference effects which have been used in preparing the plates.

ARTHUR SCHUSTER.

August, 1904.

CONTENTS

PART I

CHAPTER I

PERIODIC MOTION

CHAPTER II

KINEMATICS AND KINETICS OF WAVE MOTION

CHAPTER III

PRELIMINARY DISCUSSION OF THE NATURE OF LIGHT AND ITS PROPAGATION

CHAPTER IV

THE PRINCIPLE OF SUPERPOSITION

CHAPTER V

THE DIFFRACTION OF LIGHT

CHAPTER VI

MEASUREMENT OF WAVE-LENGTHS

CHAPTER VII

THE THEORY OF OPTICAL INSTRUMENTS

CHAPTER VIII

THE PROPAGATION OF LIGHT IN CRYSTALLINE MEDIA

CHAPTER IX

INTERFERENCE OF POLARIZED LIGHT

PART II

CHAPTER X

THEORIES OF LIGHT

CHAPTER XI

DISPERSION AND ABSORPTION

CHAPTER XII

ROTATORY EFFECTS

CHAPTER XIII

TRANSMISSION OF ENERGY

CHAPTER XIV

FURTHER DISCUSSION OF THE NATURE OF LIGHT AND ITS PROPAGATION

CHAPTER XV

EMISSION SPECTRA AND THE QUANTUM THEORY

CHAPTER XVI

DYNAMICAL THEORY OF SPECTRA

PLATES

AN INTRODUCTION TO THE THEORY OF OPTICS.

PART I.

CHAPTER I.

PERIODIC MOTION.

1. The Simple Periodic Motion. A motion which is repeated at regular intervals of time is called a periodic motion. The simplest kind of periodic motion is that in which a particle moves in a straight line, in such a way that its distance, x, from a fixed centre satisfies the equation

$$x = a \sin \omega (t - \theta) \qquad \text{.........................(1)},$$

where t is the time and a and ω are constants. The equation shows that the particle oscillates continuously between two points which are at a distance a from the centre. This distance is called the amplitude.

The velocity (u) of the particle which moves according to (1) is

$$u = a\omega \cos \omega (t - \theta) \qquad \text{.......................(2)},$$

and the acceleration (f) is

$$f = - a\omega^2 \sin \omega (t - \theta) \qquad \text{.......................(3)}.$$

The particle passes through its central position ($x = 0$) when

$$t - \theta = m\pi/\omega,$$

m being an integer. The velocity of the particle is then ωa when m is even, and $-\omega a$ when m is odd. Hence the velocity has its greatest value when $x = 0$, but may be positive or negative according as the particle passes through its central position from the negative or from the positive side.

The values of x, u and f in the above equations remain unaltered if $t + \tau$ be substituted for τ, provided that $\tau = 2m/\omega$. Hence, at regular intervals of time equal to τ, the displacement velocity and acceleration of the particle return exactly to their previous states. The time τ is called the "*time of oscillation*," "*periodic time*" or simply the "*period*" of the motion.

Equations (1) and (2) alter their form if a different constant be substituted for θ. Thus by writing $\omega\theta_1 = \omega\theta + \frac{1}{2}\pi$, we obtain

$$x = a\cos\omega(t-\theta_1) \qquad (1a),$$

$$u = -\omega a\sin\omega(t-\theta_1) \qquad (2a).$$

When dealing with one particle only, the origin of time may be chosen according to convenience, so that we may adopt the simpler forms obtained by making θ or θ_1 equal to zero in the previous equations.

I proceed to show that equations (2) and (3) are necessary consequences of (1).

In Fig. 1 consider a point P moving uniformly in a circle of radius

$$a = OA.$$

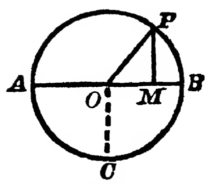


Fig. 1.

Let OM be the projection of OP on a diameter AB. If the angle POM be denoted by ϕ, and the distance OM by x,

$$x = a\cos\phi.$$

If the particle passes through the position B when $t = \theta_1$, and takes a time τ to complete a whole revolution,

$$\phi = 2\pi(t-\theta_1)/\tau.$$

Hence $$x = a\cos\omega(t-\theta_1),$$

where $$\omega = 2\pi/\tau.$$

This shows that the point M moves in the simple periodic motion indicated by equations (1a) or (1) and we have the important proposition that this periodic motion may be represented as an orthogonal projection of a uniform circular motion. The periodic time τ is the time of revolution of the point P, the amplitude is the radius of the circle, and the constant θ_1 represents the smallest positive value of the time at which the particle reaches its extreme position on the positive side.

The proper expressions for the velocity and acceleration of the point M are obtained by considering that these are equal to the projections on AB of the velocity and acceleration of P.

If the velocity of P be denoted by U:

$$u = -U\sin\phi$$
$$= -U\sin\omega(\mathrm{t}-\theta_1).$$

The minus sign is a consequence of the negative direction of the velocity of M when ϕ is positive. The whole circumference of the circle being described in a time t, it follows that

$$U = 2\pi a/\tau = a\omega.$$

Hence finally $$u = -a\omega\sin\omega(t-\theta_1).$$

The expression for the acceleration of the point M is obtained in a similar manner. The acceleration of the point P is directed radially inwards towards the centre of the circle and is equal to U^2/a, and the acceleration f of M is the projection of this acceleration upon the diameter AOB.

$$\therefore f = -(U^2 \cos \phi)/a$$
$$= -a\omega^2 \cos \omega (t - \theta_1).$$

A periodic motion may be of a more complicated character than that indicated by the above equations. If we were to take *e.g.* the orthogonal projection of a particle moving with uniform speed in an ellipse, we should get a motion which is strictly periodic, but which could not be represented by the simple equations we have given. Even the oscillations of a simple pendulum can only be *approximately* represented by our equations, the approximation being the more nearly correct, the smaller the amplitude.

I shall call a "simple" or "normal" oscillation one which can be represented as the orthogonal projection of a uniform circular motion. A normal oscillation is identical with that often called "harmonic motion." I avoid this term because "harmony" means a relation between different things, and not a property of any particular thing.

The character of the motion of a particle performing normal oscillations is completely determined by the amplitude and period, but the state of motion at any time requires a third quantity for its definition. If the oscillation is considered to be the projection of a uniform circular motion, it is convenient to take the angle between the radius vector OP (Fig. 1) and some fixed radius as the quantity defining the state of motion. This angle is called the "phase" of motion, and is to a certain extent arbitrary, as the fixed radius may be drawn in any direction.

If we express the motion in the form

$$x = a \sin \omega (t - \theta)$$

it is usual to define zero phase as the phase at the time the particle passes through its mean position in the positive direction. The radius of reference will then be OC (Fig. 1) at right angles to AB, and $\omega (t - \theta)$ will measure the phase.

On the other hand, if we choose the form

$$x = a \cos \omega (t - \theta_1)$$

for the equation of motion, we may define zero phase to be the phase at the time the particle reaches its extreme position on the positive side; then $\omega (t - \theta_1)$ will be the phase at any time during the subsequent motion, the radius of reference being OB, or the positive branch of the direction on which the motion is projected. The want of uniformity in the choice of the direction which defines the zero phase, causes no inconvenience, as we are nearly always concerned with *differences* of phase, and this difference is perfectly determinate. Thus if in Fig. 6 two periodic motions are represented by the projections of the circular motions of two particles P and Q on the same straight line, the angle POQ will always represent the difference between the phases, whatever line is taken to be the direction of zero phase.

The difference in phase between two normal periodic motions having the same period remains constant.

Representing the two motions by

$$x_1 = a_1 \cos \omega (t - \theta_1),$$
$$x_2 = a_2 \cos \omega (t - \theta_2),$$

the difference in phase will be

$$\omega (t - \theta_1) - \omega (t - \theta_2) = \omega (\theta_2 - \theta_1),$$

which proves the proposition, as t disappears in the final expression.

2. Normal oscillations under the action of forces varying as the distance. The equations for the displacement x and the acceleration f of a particle which has a simple periodic motion are

$$x = a \sin \omega (t - \theta_1),$$
$$f = - a\omega^2 \sin \omega (t - \theta_1).$$

By combining these we obtain the relation:

$$f = - \omega^2 x \quad \text{.............................(4).}$$

This is an equation of great importance, for it gives the necessary condition which must be satisfied in order that a particle may execute normal oscillations when acted on by a force directed to a centre. This condition is, that the force is proportional to the distance of the particle from the centre.

Consider a particle constrained to move in a straight line and attracted to a fixed centre by a force F, which is proportional to the displacement. If m is the mass of the particle and $F = - n^2 x$

$$f = \frac{F}{m} = - \frac{n^2}{m} x.$$

This agrees with (4) if ω^2 is equal to n^2/m, and hence τ the time of oscillation is obtained in terms of m and n, for

$$\tau = \frac{2\pi}{\omega} = \frac{2\pi\sqrt{m}}{n}$$

As all forces of nature diminish with increasing distance, the particular law of force which produces normal oscillations may not at-first sight seem to be of practical importance. It is quite true that bodies moving under the action of such central forces as occur in nature would not perform normal oscillations; but when a body, originally at rest under the action of opposing forces, is slightly displaced, the forces act so as to bring it back to its equilibrium position. The resultant of the forces on which the subsequent motion depends increases in general with its distance from the position of equilibrium, and when the displacement is small may be taken to be proportional to it. As this is an important fact, it is well to give a few illustrations.

Example 1. *The Simple Pendulum.* A heavy particle is suspended from a fixed point by a light string of length l and is set in motion.

Let θ (Fig. 2) be the angular deviation of the string from the vertical. The only forces acting on the particle are its weight and the tension of the string. The particle is constrained to move in a circle, and the force which tends to draw back the particle to its position of equilibrium is found by resolving the acting forces along the tangent to the arc.

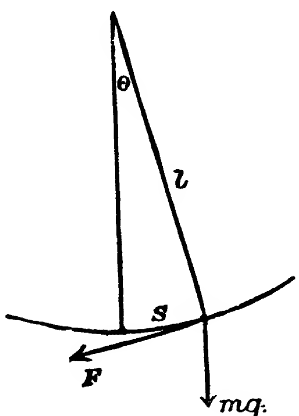


Fig. 2.

If m is the mass of the particle, its weight is mg. The tension of the string has no component in the direction of the tangent to the arc, and therefore the resultant force acting on the particle is $mg \sin \theta$.

If θ is so small that we can neglect θ^2 compared to unity, we may replace $\sin \theta$ by the angle θ, so that the force F acting on the particle is:

$$F = -mg\theta$$

$$= -mg\frac{s}{l} \quad \text{.............................(5)},$$

where s is the displacement of the particle along the arc corresponding to the angular displacement θ.

This equation shows that the particle moves along the arc, as if it were subject to a restoring force which is proportional to the distance of the particle from the lowest point of the arc. Therefore the particle will describe normal oscillations about this point. The acceleration of the particle at any distance s is F/m or $-gs/l$. By comparing this with (4) it follows that $\omega^2 = g/l$ or that the period is determined by

$$\tau = 2\pi\sqrt{\frac{l}{g}}.$$

This is the well known equation for the time of oscillation of a simple pendulum.

Example 2*. *Oscillations produced by volume elasticity of gaseous pressure.*

Let an airtight vessel be closed by a weighted piston A, which can move without friction in a cylindrical tube attached to the vessel. This piston will have a definite position of equilibrium. If it is forced down below this position and then released, it will be driven up again by the increased pressure of the air within the vessel. The momentum it then acquires will carry it past its position of equilibrium. The air in the vessel expands to fill the larger volume, its pressure is accordingly reduced, and it is unable to counterbalance the weight of the piston and the external pressure. The piston is thus driven in again, and, the process repeating itself at regular intervals, periodic oscillations are performed.

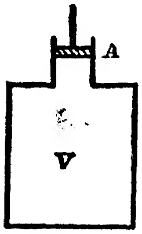


Fig. 3.

We proceed to find the time of oscillation of the piston. Let V be the original volume of the vessel and P the pressure of the enclosed air. Suppose the piston is pushed down until the volume is diminished by a small quantity v and the pressure is increased by a small amount p. The volume and pressure are then $(V-v)$ and $(P+p)$ respectively. We shall disregard the inertia of the air and assume the motion to be sufficiently slow to allow the change to be isothermal. We have then, applying Boyle's Law:

$$VP=(V-v)(P+p)$$

or

$$pv=pV-vP.$$

If the displacements are small, so that the product of the two small quantities p and v may be neglected,

$$pV=vP$$

or

$$p=P\frac{v}{V}.$$

Denote by A the area of the base of the piston. Then the resultant force on the piston, when the volume of the vessel is diminished by v, is

$$\begin{aligned}F&=pA\\&=\frac{AP}{V}.v.\end{aligned}$$

If x is the distance through which the piston moves

$$v=-A.x$$

and

$$F=-\frac{A^2P}{V}.x.$$

* Examples 2 and 3 are taken from Lord Rayleigh's *Sound* where they are treated in a different manner.

Thus the force acting on the piston is proportional to the displacement of the piston from its equilibrium position and is always in the opposite direction to the displacement. If M is the mass of the piston, its acceleration is F/M or $\frac{A^2P}{MV}\cdot x$. Hence the time of oscillation is given by

$$\tau = 2\pi\sqrt{\frac{MV}{A^2P}}.$$

If the vessel were a cylinder of length l and area A, so that $V=Al$, and if also the pressure were entirely due to the weight Mg of the piston, we should have

$$P=\frac{Mg}{A},$$

and by substitution it would follow that

$$\tau = 2\pi\sqrt{\frac{l}{g}},$$

i.e. the time of oscillation of the piston would be exactly the same as the time of oscillation of a simple pendulum, the length of which is the same as that of the cylindrical vessel.

Example 3. *Normal Oscillations due to the tension of a string.*

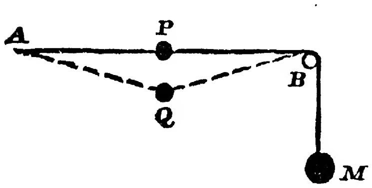


Fig. 4.

Let a string attached to A pass over a peg B at the same level as A, and carry at its end a mass M. If a particle P of mass m be attached to the string halfway between A and B, and the particle be displaced vertically downwards until it coincides with Q, the tension T of the string will have a resultant vertically upwards which, neglecting the weight of the string, is easily shown to be

$$-\frac{2Tx}{\sqrt{a^2+x^2}},$$

where $2a$ is the distance AB and x the displacement PQ.

If x is so small that x^2/a^2 may be neglected, and if m is so small compared to M that in the position of equilibrium the displacement of P is a small quantity of the second order, we may disregard the weight of P in calculating the tension which will then be equal to Mg. The acceleration of P is therefore $-\frac{2Mgx}{am}$, and hence the time of oscillation

$$\tau = 2\pi\sqrt{\frac{am}{2Mg}}.$$

3. Energy of a Particle in Periodic Motion. If a particle whose mass is m is moving with a velocity v, its kinetic energy is $\frac{1}{2}mv^2$.

If the particle is executing normal oscillations, its velocity at any time t is

$$u = a\omega \cos \omega (t - \theta).$$

Therefore its kinetic energy E at this instant is

$$\begin{aligned} E &= \tfrac{1}{2} m\omega^2 a^2 \cos^2 \omega (t - \theta) \\ &= \tfrac{1}{4} m\omega^2 a^2 \{1 + \cos 2\omega (t - \theta)\}. \end{aligned}$$

The second term on the right-hand side has values ranging from $+1$ to -1, and is as often positive as negative, its average value taken over one complete period of vibration being zero.

If $U = a\omega$ is the maximum velocity of the particle, it follows that the average value of E is $\tfrac{1}{4} ma^2\omega^2$ or $\tfrac{1}{4} mU^2$. This proves that the *average* energy is half the *maximum* energy.

The average value of the kinetic energy of a vibrating particle is taken as the measure of the intensity of the vibration, which has just been shown to be proportional to the square of the amplitude as long as the mass and period remain the same.

In the simple cases we have been considering no energy leaves the particle, and hence all changes of kinetic energy must be compensated by corresponding changes in the potential energy. Now the kinetic energy E varies, being at its maximum of $\tfrac{1}{2} mU^2$ when the particle is passing through its central position and falling to zero when the displacement is a maximum. If the constancy of the total energy is maintained, it follows that the potential energy P must satisfy the equation

$$P + \tfrac{1}{2} mu^2 = \text{a constant}.$$

Assuming the potential energy to be zero when the kinetic energy is at its maximum, the value of the constant must be $\tfrac{1}{2} mU^2$. Hence

$$\begin{aligned} P &= \tfrac{1}{2} m (U^2 - u^2) \\ &= \tfrac{1}{2} ma^2\omega^2 \{1 - \cos^2 \omega (t - \theta)\} \\ &= \tfrac{1}{2} ma^2\omega^2 \sin^2 \omega (t - \theta) \\ &= \tfrac{1}{2} m\omega^2 x^2. \end{aligned}$$

Thus for a body performing normal oscillations the potential energy is proportional to the square of the displacement of the particle from the centre of force. This proposition may be proved independently from the work done by the forces.

4. Composition of Periodic Motions. If a single particle is acted upon by two distinct agents, each of which, if acting separately, would cause the particle to perform simple periodic vibrations, the question arises—What is the resultant motion on the supposition that each produces its own effect?

We consider first the case in which the two component vibrations are in the same straight line and have the same period.

Let the two amplitudes be a_1 and a_2 and the common period, $2\pi/\omega$. At any instant t, the displacement x_1, due to the first oscillation, would be

$$x_1 = a_1 \cos \omega (t - \theta_1)$$

and that due to the second oscillation

$$x_2 = a_2 \cos \omega (t - \theta_2).$$

Since the two displacements are in the same straight line and each produces its own effect, we can combine them algebraically and write for the resultant displacement

$$\begin{aligned} x &= x_1 + x_2 \\ &= a_1 \cos \omega (t - \theta_1) + a_2 \cos \omega (t - \theta_2) \\ &= P \cos \omega t + Q \sin \omega t, \end{aligned}$$

where

$$\left.\begin{aligned} P &= a_1 \cos \omega\theta_1 + a_2 \cos \omega\theta_2 \\ Q &= a_1 \sin \omega\theta_1 + a_2 \sin \omega\theta_2 \end{aligned}\right\} \quad \text{..........................(6).}$$

Now write

$$\begin{aligned} P &= R \cos \delta \\ Q &= R \sin \delta, \end{aligned}$$

so that

$$R^2 = P^2 + Q^2 \quad \text{and} \quad \tan \delta = Q/P.$$

Then

$$\begin{aligned} x &= R (\cos \omega t \cos \delta + \sin \omega t \sin \delta) \\ &= R \cos (\omega t - \delta) \quad \text{....................................(7).} \end{aligned}$$

It is seen from the last equation that the two component simple periodic oscillations have combined to form a resultant simple periodic oscillation with the same period as the component oscillations, but with a different amplitude and phase.

The amplitude of the resultant oscillation is R,

where

$$\begin{aligned} R^2 &= P^2 + Q^2 \\ &= (a_1 \cos \omega\theta_1 + a_2 \cos \omega\theta_2)^2 + (a_1 \sin \omega\theta_1 + a_2 \sin \omega\theta_2)^2 \\ &= a_1^2 + a_2^2 + 2a_1 a_2 \cos \omega (\theta_2 - \theta_1). \end{aligned}$$

Therefore the amplitude $R = \sqrt{\{a_1^2 + a_2^2 + 2a_1 a_2 \cos \omega (\theta_2 - \theta_1)\}}$ is equal in magnitude to the diagonal of a parallelogram having two adjoining sides a_1 and a_2, the angle between the sides being $\omega (\theta_2 - \theta_1)$.

We may now show that the diagonal OP not only represents the resultant oscillation in amplitude, but also indicates the phase.

By separating the quantities a_1 and a_2 in the equations (6) we obtain the two equations:

$$\begin{aligned} P \sin \omega\theta_2 - Q \cos \omega\theta_2 &= a_1 \sin \omega (\theta_2 - \theta_1), \\ P \sin \omega\theta_1 - Q \cos \omega\theta_1 &= -a_2 \sin \omega (\theta_2 - \theta_1). \end{aligned}$$

If in these we substitute $P = R \cos \delta$ and $Q = R \sin \delta$, we find

$$\begin{aligned} R \sin (\omega\theta_2 - \delta) &= a_1 \sin \omega (\theta_2 - \theta_1), \\ R \sin (\omega\theta_1 - \delta) &= -a_2 \sin \omega (\theta_2 - \theta_1), \end{aligned}$$

or, expressed otherwise

$$a_1 : a_2 : R = \sin(\omega\theta_2 - \delta) : \sin(\delta - \omega\theta_1) : \sin\omega(\theta_2 - \theta_1) \quad \ldots(8).$$

If a parallelogram be described with $OA = a_1$, and $OB = a_2$, as adjoining sides, and with an angle AOB equal to the difference in phase $\omega(\theta_2 - \theta_1)$, between the two oscillations which are to be combined, the geometry of Fig. 5 shows that

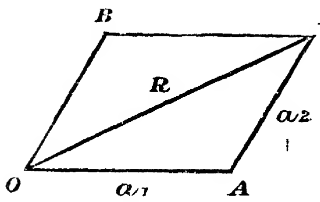


Fig. 5.

$$a_1 : a_2 : R = \sin BOP : \sin AOP : \sin AOB.$$

A comparison between this relation and (8) shows that

$$BOP = \omega\theta_2 - \delta \quad \text{and} \quad AOP = \delta - \omega\theta_1,$$

which means that the angles between the diagonal OP and the two lines a_1 and a_2 represent the difference in phase between the resultant oscillation and the two component oscillations.

The proposition that normal oscillations in the same straight line may be combined like two forces is of primary importance, and a second proof of it is therefore given.

We represent the two periodic motions by the orthogonal projections OM_1 and OM_2 of two points P and Q, moving with uniform speeds round two circles (Fig. 6).

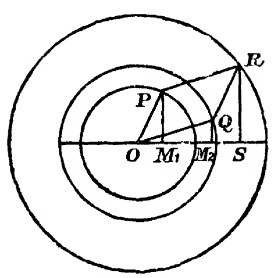


Fig. 6.

The two radii OP, OQ represent the amplitudes of the respective oscillations. If the periodic time is the same the radii OP, OQ revolve with the same angular velocity, and therefore the angle POQ remains constant. Complete the parallelogram $OQRP$ and imagine a third point at the angle R. Then the point at R will describe a circle in the same time as the points P and Q, and its projection S on the diameter AB will perform a simple periodic vibration.

But $$OS = OM_1 + OM_2$$

since the projection of OR must equal the sum of the projections of OP and OQ.

Hence the displacement of S is always equal to the sum of the displacements of M_1 and M_2 and the motion of S will be the resultant of the motions of M_1 and M_2.

The figure shows that the resultant amplitude OR is found from the amplitudes OP, OQ by the parallelogram construction and that this construction enables us to determine not only the amplitude but also the phase of the resultant motion. For, if we measure phases from the direction determined by the maximum positive displacement, then in the figure, the phase of the particle M_1 is the angle POS and the phases of M_2 and S are measured by the angles QOS and ROS respectively.

Hence the direction of the diagonal OR indicates the phase of the resultant oscillation.

5. Combination of any number of Oscillations. Having seen how two linear oscillations, which are in the same straight line, can be combined, it follows that any number of such oscillations can be combined by taking the resultant of any two of them, and combining with it a third oscillation and so on, until we reach the final resultant. In short, a system of such oscillations is reduced to a single resultant in exactly the same way as a system of forces acting at a point. Any proposition relating to a system of forces can be made to apply to a system of linear oscillations of the same period which take place in the same straight line.

According to a well known proposition in Statics a system of n forces OP_1, OP_2 ......OP_n has a resultant which coincides in direction with OG and is in magnitude equal to nOG, if G is the centre of inertia of particles having equal masses, placed at points P_1, P_2......P_n. We make use of this proposition to find the resultant of a large number of oscillations of equal amplitude and having their phases in arithmetic progression.

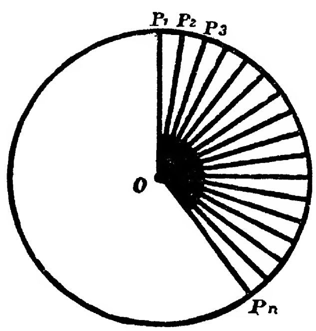


Fig. 7.

The oscillations will be represented by the lines OP_1, OP_2, OP_3, ... OP_n (Fig. 7), such that all the points P are equidistant and lie on the arc of a circle. If the constant phase difference between two successive oscillations is very small, the problem of finding the resultant resolves itself into the determination of G, the position of the centre of inertia of the arc P_1P_n.

The distance OG is known to be equal to $\dfrac{a \sin \alpha}{\alpha}$ where 2α is the angle of the arc P_1P_n and hence α is the angle between OG and either OP_1 or OP_n. The resultant vibration has therefore an amplitude equal to $na \sin \alpha/\alpha$ and a phase which lies halfway between the phases of the first and last vibrations. If all vibrations were of equal phase, the resultant amplitude would be na.

Hence we may formulate the following important proposition proved by the above reasoning:

Normal rectilinear oscillations which have equal amplitudes and periods, and take place along the same straight line such that any two successive oscillations have a phase difference which is small and equal for each successive pair, combine together into a resultant oscillation which has the same period, and a phase halfway between that of the first and last oscillation. The amplitude of the resultant oscillation is $R \sin \alpha/\alpha$ where 2α is the phase difference between the two extreme oscillations and R the amplitude of the resultant in the special case that the phase differences vanish. The values of $\sin \alpha/\alpha$ and $\sin^2 \alpha/\alpha^2$ are plotted as ordinates as against α as abscissa in Fig. 70, Art. 54.

6. Combination of oscillations in directions at right angles to each other. Let a particle M (Fig. 8) describe simple periodic oscillations in the direction OX about the centre O, its motion being represented by the equation

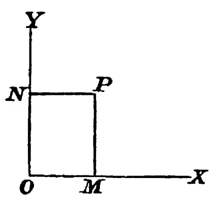


Fig. 8.

$$x = a_1 \cos \omega t.$$

Also let a second particle N perform oscillations of the same period, about the same centre O, but in the direction OY perpendicular to OX. The motion of N may be represented by the equation

$$y = a_2 \cos (\omega t + \delta),$$

δ expressing the difference of phase between the two oscillations. Now imagine a third particle P to move in such a way that its projections on OX and OY always coincide with the points M and N. The problem is to investigate the motion of the particle P. Before treating the question generally we may take a few cases, which are simple and of special importance.

Case I. Let $\delta = 0$. This means that both M and N pass through the centre O at the same instant, and that therefore the point P passes through O.

The equations of motion of M and N are respectively:

$$x = a_1 \cos \omega t,$$
$$y = a_2 \cos \omega t.$$

By eliminating the time t from the equations, we obtain a relation between x and y, which determines the path described by P.

Thus $$\frac{x}{a_1} = \frac{y}{a_2} \quad \text{or} \quad y = \frac{a_2}{a_1} x.$$

This is the equation of a straight line passing through the origin O. The cosines of the angles which OP forms with OX and OY respectively are $a_1/\sqrt{a_1^2 + a_2^2}$ and $a_2/\sqrt{a_1^2 + a_2^2}$. Projecting x and y on OP we see that the distance (r) of P from the origin is $r = \sqrt{a_1^2 + a_2^2} \cos \omega t$. Therefore the motion of the particle P is a simply periodic linear oscillation in the direction OP, having the same periodic time as its component vibrations, and an amplitude equal to $\sqrt{a_1^2 + a_2^2}$.

Case II. *Let* $\delta = \pm\frac{\pi}{2}$. This means that the particle M is passing through its mean position when N has its maximum displacement.

The equations of motion are now:

$$\left.\begin{aligned} x &= a_1 \cos \omega t \\ y &= a_2 \cos\left(\omega t \pm \frac{\pi}{2}\right) \end{aligned}\right\} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(9)$$

$$= \mp a_2 \sin \omega t.$$

Eliminating t by squaring and adding the two equations, it is found that

$$\frac{x^2}{a_1^2} + \frac{y^2}{a_2^2} = 1.$$

Hence the path described by the particle P is an ellipse. In the special case of $a_1 = a_2$ the equation becomes that of a circle of radius a_1.

The time occupied by P in moving round the ellipse or circle is the same as the periodic time of the linear vibrations. It is easily seen that if the phase of M moving along OX *exceeds* by a right angle the phase of N moving along OY, the motion will be from the positive axis of x to the positive axis of y, for according to equations (9) the particle P crosses the positive axis of x when $t=0$. When $t = \pi/2\omega$ or after a quarter of a period, P is on the negative or positive branch of OY according as the upper or lower sign is taken. Hence the positive sign in the second equation (9) indicates a clockwise and the negative sign an anti-clockwise revolution. The axes of the ellipse in which P moves are coincident with the axes of x and y.

Case III. or General Case. Let δ now have any value whatever. The equations of motion are

$$\left.\begin{aligned} x &= a_1 \cos \omega t \\ y &= a_2 \cos(\omega t + \delta) \end{aligned}\right\} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(10).$$

Then
$$\frac{y}{a_2} = \cos \omega t \cos \delta - \sin \omega t \sin \delta$$

$$= \frac{x}{a_1} \cos \delta - \sin \omega t \sin \delta,$$

$$\therefore \sin \omega t \sin \delta = \frac{x}{a_1} \cos \delta - \frac{y}{a_2},$$

and from (10):
$$\cos \omega t \sin \delta = \frac{x}{a_1} \sin \delta.$$

Squaring and adding we get

$$\sin^2 \delta = \frac{x^2}{a_1^2} - 2\frac{xy}{a_1 a_2} \cos \delta + \frac{y^2}{a_2^2} \quad \ldots\ldots\ldots\ldots\ldots\ldots(11).$$

This is the equation of an ellipse the axes of which are not now, in general, parallel to the x and y axes (Fig. 9). By putting $\delta=0$ or $\delta=\pm\frac{\pi}{2}$ in the general equation we return to the special cases I. and II.

In the customary treatment of the kinematics of a particle, the point P is said to possess simultaneously velocities along OX and along OY which are respectively equal to the velocities of its projections. Adopting this mode of expression, we may say that a particle having two simple periodic motions of equal period at right angles to each other, moves in general in an ellipse, the time of revolution being equal to the period of the oscillation. In special cases the ellipse may become a circle or a straight line.

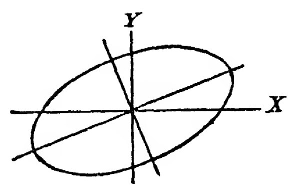


Fig. 9.

If two periodic motions at right angles to each other be combined, we may determine the direction and magnitude of the principal axes of the resulting elliptic orbit. We introduce for this purpose a second system of coordinates inclined at an angle γ to the original one. The new and old coordinates are connected by

$$\left.\begin{aligned} x &= x'\cos\gamma - y'\sin\gamma \\ y &= x'\sin\gamma + y'\cos\gamma \end{aligned}\right\} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots(12).$$

Substituting these expressions into (11) it is found that the factor of $x'y'$ is zero, when

$$\tan 2\gamma = \frac{2a_1a_2}{a_1^2-a_2^2}\cos\delta,$$

or introducing an auxiliary angle ψ defined by $\tan\psi = a_2/a_1$ the relation becomes

$$\tan 2\gamma = \tan 2\psi\cos\delta \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots(13).$$

The vanishing of the factor of $x'y'$ implies that the ellipse is now referred to its principal axes and γ therefore is the angle which these principal axes form with the original directions of vibration. The magnitude of the semi-axes is found in a similar manner.

In investigating elliptic polarisation of light the converse problem sometimes presents itself: the ratio of the principal axes of an elliptic orbit and their inclination to fixed directions being found by experiment, we may require to calculate the ratio of amplitudes and the relative phase of the two normal vibrations in the fixed directions into which the elliptic motion may be resolved.

To solve this problem we make use of two well-known propositions of analytical geometry relating to the equation of a conic:

$$px^2+qy^2+2cxy=1 \ldots\ldots\ldots\ldots\ldots\ldots\ldots(14),$$

which, transformed by turning the axes of reference through an angle γ, becomes

$$Px'^2 + Qy'^2 + 2Cx'y' = 1.$$

The two propositions referred to are then expressed by

$$\left.\begin{aligned} P + Q &= p + q \\ \text{and} \qquad PQ - C^2 &= pq - c^2 \end{aligned}\right\} \quad \text{..........(15).}$$

If in the transformed equation the ellipse be referred to its principal axes $C = 0$, and we may then proceed as follows:

$$\frac{P+Q}{\sqrt{PQ}} = \frac{p+q}{\sqrt{pq-c^2}} \quad \text{..........(16).}$$

We now substitute the values of p, q and c which are found by com paring (11) and (14) and introduce an auxiliary angle Ψ, defined by $\tan\Psi = B/A$ where A and B are the semi-axes of the ellipse and therefore equal to $P^{-\frac{1}{2}}$ and $Q^{-\frac{1}{2}}$ respectively. Equation (16) then becomes

$$\sin 2\Psi = \sin 2\psi \sin \delta \quad \text{..........(17).}$$

In the present problem Ψ and γ are known, while ψ and δ are to be determined. Equations (17) and (13) suffice for the purpose. To separate the variables we derive from (17)

$$\cos 2\Psi = \cos 2\psi \sqrt{1 + \cos^2\delta \tan^2 2\psi}$$

which with the help of (13) becomes

$$\cos 2\Psi \cos 2\gamma = \cos 2\psi \quad \text{..........(18).}$$

This determines ψ.

Combining the last equation with (13) we find

$$\cos 2\Psi \sin 2\gamma = \sin 2\psi \cos \delta \quad \text{..........(19).}$$

Finally the division of (17) by (19) gives

$$\tan 2\Psi = \sin 2\gamma \tan \delta \quad \text{..........(20),}$$

which determines δ.

From (10) we obtain for the components of velocity u and v

$$u^2 = \omega^2 (a_1^2 - x^2),$$
$$v^2 = \omega^2 (a_2^2 - y^2).$$

Hence for the velocity in the elliptic orbit

$$\begin{aligned} U^2 &= u^2 + v^2 \\ &= \omega^2 (a_1^2 + a_2^2 - r^2) \quad \text{..........(21),} \end{aligned}$$

where r is the distance of the moving point from the origin. We conclude that the motion of a point in an elliptic orbit can only be represented by the superposition of two periodic motions at right angles to each other if the velocity in the orbit follows a perfectly definite law. If that law is satisfied the motion resolved along any axis is simply periodic.

It has been shown that a uniform circular motion may be resolved into two simple periodic motions at right angles to each other.

Conversely it may easily be proved either geometrically or algebraically that a simple periodic motion may be resolved into two circular motions of equal amplitudes and opposite velocities.

Any number of simple periodic motions in a plane, having the same period but differing in amplitude and phase, may be combined into an elliptic motion. This follows at once because a periodic oscillation may be decomposed into two along the same fixed axes at right angles to each other. Adding the components which lie in the same direction according to Art. 4 and then combining the two resultant oscillations at right angles to each other we obtain the resulting elliptic motion.

Any number of simple periodic motions in a plane, having the same period but differing in amplitude and phase, may be combined into two uniform circular motions in opposite directions, but not necessarily along circles of equal radii. This must be true because each of them may be decomposed into two opposite circular motions, and all circular motions having the same direction may be combined again into a uniform circular motion.

It follows that any elliptic motion in which the velocity satisfies the condition (21), may be considered as being composed of two uniform circular motions in opposite directions.

To prove this algebraically, let the rectangular projections of one circular motion taking place anti-clockwise be $a_1 \cos \omega t$ and $a_1 \sin \omega t$ and that of another circular motion taking place clockwise $a_2 \cos \omega (t-\theta)$ and $-a_2 \sin \omega (t-\theta)$ so that their combined motion is represented by

$$x = a_1 \cos \omega t + a_2 \cos \omega (t-\theta),$$
$$y = a_1 \sin \omega t - a_2 \sin \omega (t-\theta).$$

Eliminating t in the usual way, gives for the elliptic path the quadratic equation

$$x^2 (a_1^2 + a_2^2 - 2a_1a_2 \cos \omega\theta) + y^2 (a_1^2 + a_2^2 + 2a_1a_2 \cos \omega\theta) - 4xya_1a_2 \sin \omega\theta = (a_1^2 - a_2^2)^2.$$

The three available constants a_1, a_2, and θ may now be determined in terms of the three constants which determine the elliptic orbit.

7. Composition of Linear Vibrations of slightly different Periodic Times. We now consider the composition of two linear vibrations in the same direction but having slightly different periodic times.

Let the displacements be represented by

$$x_1 = a \cos \omega_1 t,$$
$$x_2 = a \cos \omega_2 t,$$

assuming, for simplicity, that they have the same amplitude. The resultant vibration is given by

$$x = x_1 + x_2 = a \cos \omega_1 t + a \cos \omega_2 t$$
$$= 2a \cos \frac{\omega_1 + \omega_2}{2} t \,.\cos \frac{\omega_1 - \omega_2}{2} t.$$

The factor $\cos \frac{\omega_1 - \omega_2}{2} t$ is periodic, varying in value between $+1$ and -1 and going through a complete period in the time $4\pi/(\omega_1 - \omega_2)$. Now this time is great (because $\omega_1 - \omega_2$ is small) in comparison with the time $4\pi/(\omega_1 + \omega_2)$ which is the period of the other factor. We may therefore consider $2a \cos \frac{\omega_1 - \omega_2}{2} t$ to be the slowly varying amplitude of a simple oscillation, having a period $4\pi/(\omega_1 + \omega_2)$.

The intensity I of the resultant vibration is proportional to the square of the amplitude, so that

$$I \propto 4a^2 \cos^2 \frac{\omega_1 - \omega_2}{2} t$$
$$\propto 2a^2 \{1 + \cos(\omega_1 - \omega_2) t\}.$$

Hence the resultant intensity varies between $4a^2$ and 0, and the time interval between two successive maxima of intensity is $2\pi/(\omega_1 - \omega_2)$. An important application of this equation is made in the theory of sound. When two notes of nearly equal pitch are sounded together, beats are heard, and according to the above, the periodicity of the beats is $2\pi/(\omega_1 - \omega_2)$, if $2\pi/\omega_1$ and $2\pi/\omega_2$ are the periods of the two notes. As the number of vibrations per second (the frequencies) are inversely proportional to the periods, it follows that when two notes have frequencies n_1 and n_2, the number of beats per second is $n_1 - n_2$.

8. Use of imaginary quantities. The mathematical treatment of oscillations may often be made more concise by the introduction of imaginary quantities. Writing $i = \sqrt{-1}$, we make use of the symbolic expression

$$e^{i\phi} = \cos \phi + i \sin \phi.$$

If $\phi = \omega t$, it is seen that both the real and imaginary part of $e^{i\phi}$, represents a simple periodic motion. The same is true for $ce^{i\phi}$, where the "amplitude" c, may be real, imaginary, or complex. Writing, to separate the real and imaginary parts, $c = a + ib$, and

$$\left.\begin{array}{l} r \cos \delta = a \\ r \sin \delta = b \end{array}\right\} \quad \text{..........................(22)},$$

it follows that $\qquad c = re^{i\delta}$ and $ce^{i\phi} = re^{i(\phi + \delta)}$.

This represents a normal oscillation of amplitude r, equal to $\sqrt{a^2 + b^2}$ and having a phase δ determined by

$$\tan \delta = b/a.$$

If the factor of $e^{i\phi}$ has the form

$$c = \frac{a + ib}{A + iB}$$

the fraction is reduced to the standard form by multiplying its numerator and denominator by $A - iB$.

We derive $$c = \frac{(aA + bB) + i(bA - aB)}{A^2 + B^2}.$$

The amplitude and phase of the real part of $ce^{i\phi}$ are now obtained from:

$$\left.\begin{aligned} r^2 &= \frac{a^2 + b^2}{A^2 + B^2} \\ \tan\delta &= \frac{bA - aB}{aA + bB} \end{aligned}\right\} \quad \text{..........................(23).}$$

For the particular case that $A = a$ and $B = -b$,

$$c = \frac{(a^2 - b^2) + 2iab}{a^2 + b^2},$$

and $$r = 1\ ;\ \tan\delta = \frac{2ab}{a^2 - b^2}.$$

According to the above, an expression of the form $\frac{a + ib}{A + iB}$ can always be brought to the form $re^{i\delta}$ where r and δ satisfy equations (23).

The transformation by which a fraction containing an imaginary part in both the numerator and denominator may be expressed in the form $re^{-\delta}$ is conveniently conducted as follows:

If the expression to be reduced is

$$P = \frac{a + ib}{A + iB};\ \text{ write }\ Q = \frac{a - ib}{A - iB}$$

when Q is obtained from P by changing the sign of i. By comparison with (23) it may then be easily shown that

$$PQ = r^2 \quad \text{and} \quad \frac{P - Q}{P + Q} = i\tan\delta.$$

CHAPTER II.

KINEMATICS AND KINETICS OF WAVE MOTION.

9. Kinematics of Wave Motion. Every one is familiar with the appearance of a train of waves propagated over a surface of water. As a rule, such surface waves alter their shape as they proceed and they are not therefore very good examples of simple wave propagation. We say that a wave has "constant type" when the outline of the wave always remains the same. Waves of sound and waves of light propagated through a vacuum are waves of constant type.

Consider a row of particles lying on a straight line, which we shall take to be the axis of x. Let the particles be displaced in a direction at right angles to x, the displacement being represented by the equation $y=f(x)$.

If the displacements at each point alter in such a way that a line joining all the particles seems to travel with velocity v in the positive direction without change of shape, the equation of the outline at any time t may still be represented by the same equation $y=f(x)$, provided the origin from which x is measured is shifted through a distance vt. Referred to the old origin, the equation representing the outline will be given by $y=f(x-vt)$. This then is the general equation of a wave of constant type propagated in the positive direction with a velocity v, and every wave propagated without change of shape must be expressible in this form. The argument does not turn upon the displacement y being necessarily at right angles to the direction of propagation, but it may be, as in the case of sound waves, along that direction, and the equation would hold equally for displacements of any kind. By giving to v a negative sign, we obtain the general equation of a wave propagated along x in the negative direction.

Fig. 10.

As an example we may consider the equation

$$y = ae^{-(x-vt)^2}$$

which being of the form $y=f(x-vt)$ represents a wave motion.

Putting $t=0$ we obtain for the shape of the wave, the outline

$$y = ae^{-x^2}.$$

The equation represents therefore a wave of the form shown in Fig. 10 propagated with a constant velocity v in the positive direction.

Returning to the general equation

$$y = f(x - vt)$$

we obtain by differentiation

$$\frac{dy}{dt} = -vf',$$

$$\frac{dy}{dx} = f';$$

$$\therefore \frac{dy}{dt} = -v\frac{dy}{dx} \quad \text{.............................(1).}$$

Also

$$\frac{d^2y}{dt^2} = v^2 f'',$$

$$\frac{d^2y}{dx^2} = f'';$$

$$\therefore \frac{d^2y}{dt^2} = v^2\frac{d^2y}{dx^2} \quad \text{.............................(2).}$$

The last equation is the differential equation which characterises a wave motion. Its complete solution is

$$y = f(x - vt) + F(x + vt)$$

where f and F are arbitrary functions.

As an important special case we take

$$y = a\cos(\omega t - px) \quad \text{.......................(3).}$$

By comparison with the general expression, it is seen that ω/p is the velocity of propagation. If y is measured at right angles to x, and if each point always keeps the same distance from the plane $x=0$, the motion will be rectilinear. For a given x the equation is of the form

$$y = a\cos(\omega t + \delta)$$

and every point therefore performs normal oscillations having a period

$$2\pi/\omega.$$

The outline of the wave is obtained by taking any value of t, *e.g.* $t=0$, when

$$y = a\cos px$$

will represent the shape of the wave and its position at that time. A portion of the wave form which reaches out to infinity in both directions, is represented in Fig. 11. The figure illustrates the method of drawing the curve. Equidistant points divide the circumference of a circle into equal portions. In the figure that number is twelve, but could be increased if it is desired to obtain a greater number of points in the curve. Other equidistant points are taken on a straight line

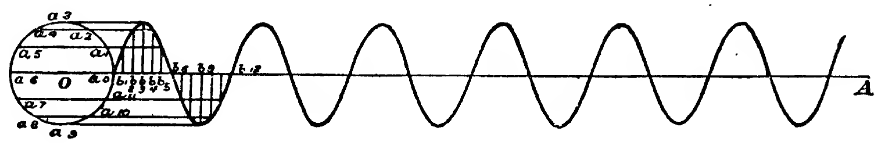


Fig. 11.

OA passing through the centre of the circle. Drawing perpendiculars to OA through each point on that line, and lines parallel to OA through the corresponding points of the circle, the intersections of the two sets of lines mark the points on the curve. The *wave-length* is the distance between the two nearest points which have the same phase. If λ be the wave-length, so that the phase is the same for x and $x+\lambda$, it follows that $p\lambda$ must be equal to 2π, or $p=2\pi/\lambda$.

From $v=\omega/p$ and $\omega=2\pi/\tau$, we obtain $v=\lambda/\tau$.

In terms of λ and τ we may write equation (3)

$$y = a\cos 2\pi\left(\frac{t}{\tau}-\frac{x}{\lambda}\right).$$

The difference of phase between two particles at distances x_1 and x_2 from the origin, as obtained from this equation, is

$$\frac{2\pi}{\lambda}(x_2-x_1).$$

In the further consideration of wave motion, we shall consider principally waves the displacement of which can be represented by the equation (3).

10. Application of Fourier's Theorem. By an important theorem due to Fourier, any function $f(x)$ may between fixed limits $x=-c$ and $x=+c$ be represented as the sum of a series, in such a way that writing $\alpha=\pi x/c$

$$f(x) = a_0 + a_1\cos\alpha + a_2\cos 2\alpha + a_3\cos 3\alpha + \dots\dots$$
$$+ b_1\sin\alpha + b_2\sin 2\alpha + b_3\sin 3\alpha + \dots\dots\dots\dots\dots\dots\dots\dots (4).$$

The constants a_0, a_1, b_1, b_2, etc. may be determined from the function f, and we may for our present purpose fix for $f(x)$ outside the specified limits the values calculated from the series on the right-hand side. If waves of all lengths are propagated with the same velocity v, we may obtain the shape at any subsequent time for waves travelling in the positive direction by writing in all terms on the right-hand side $x-vt$ for x, and having done so we may add the series again, when it is seen that the sum now becomes $f(x-vt)$. Hence the condition that normal waves of all lengths travel with the same velocity carries with it the consequence that waves of any shape may be propagated without change of type. On the other hand, if as in the case of light-waves travelling through a dispersive medium, the velocity of propagation depends on the wave-length, there must always be a change of type when waves which are not of the simple cosine or sine shape are propagated.

11. Waves travelling along a stretched string. Let us now consider the kinetics of wave propagation.

Consider a small portion AB of a curved string which is acted on by equal tangential forces at the ends. The resultant of the forces bisects AB and passes through C the centre of the circle of curvature of AB. If 2θ be the angle subtended by AB at C, the intensity of the resultant is $2T\sin\theta$, or $2T\theta$ if θ be sufficiently small. As $2r\theta$ is the length of the arc AB, if r is the radius of curvature, the "resultant force per unit length" acting on AB is T/r, *i.e.* equal to the product of the tension and the curvature.

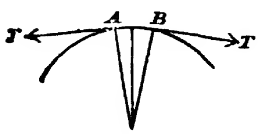


Fig. 12.

Let now a string be only slightly curved, so that every part of it lies near a straight line which shall be the axis of x. The tangent of the angle of inclination (dy/dx) may be supposed to be sufficiently small to allow its square to be neglected. The force acting on an element ds at right angles to its length has been proved to be Tds/r, and we may take the same expression to represent that component of the force which lies in the y direction.

If ρ be the mass per unit length, and hence ρds the mass of the length ds, the equation of motion is :

$$\rho ds\frac{d^2y}{dt^2}=\frac{Tds}{r};$$

$$\therefore\ \frac{d^2y}{dt^2}=\frac{T}{\rho}\cdot\frac{1}{r}.$$

Again neglecting $\left(\frac{dy}{dx}\right)^2$, the curvature is equal to $\frac{d^2y}{dx^2}$, hence

$$\frac{d^2y}{dt^2}=\frac{T}{\rho}\frac{d^2y}{dx^2}\ \ldots\ldots\ldots\ldots\ldots\ldots(5).$$

Comparing this equation with (2) it is seen that $\sqrt{T/\rho}$ is the velocity of the waves which are propagated along the string.

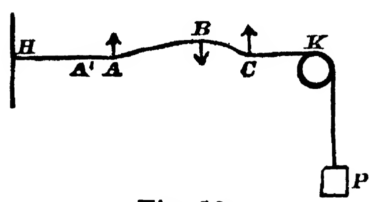


Fig. 13.

Let Fig. 13 represent a portion of a string stretched to a constant tension by a weight P. Let it be displaced by outside forces until it occupies a position such as that shown in the figure. If the constraint is suddenly removed, the tension of the string will, by what precedes, act in such a way that there is at each point a resultant force towards the centre of curvature. Hence the point B will begin to move downwards while A and C move upwards. If AH has been previously straight, this portion of the string is in equilibrium, but as soon as A is lifted up, the point at which the straight and curved portions join, has been moved to the left. If A' is that point, AA' which was previously in equilibrium, has ceased to be so. It follows that a disturbance will set out from A and travel from right to left, with a velocity which has already been found to be $\sqrt{T/\rho}$. A similar reasoning shows that the displaced region AC will also send out a disturbance from C towards K. Two waves travelling in opposite directions will therefore start from ABC.

Now we know from observation that it is possible for a disturbance to travel in one direction only, and it is a matter of interest to examine the conditions under which a displacement such as ABC may be propagated forward only or backward only. In order that it shall travel only forward, it is clearly necessary that the point A should remain in its position in spite of the force acting upwards, and this is only possible if at the time to which the figure applies, A has a velocity downwards, of such magnitude that the force acting at A just destroys the velocity. This force is of the nature of an "impulse" because if there is a discontinuity of slope at A, the curvature is infinite, and hence the force is infinite, and capable of suddenly destroying a finite velocity. Similarly all along ABC a certain relation between velocity and slope must hold, and this relation must be of such a nature that each portion will have zero velocity as soon as the wave has passed over it. The mathematical relation which must connect the slope and the velocity at each point when waves are propagated in one direction only, is obtained from (1) substituting the value of v:

$$\frac{dy}{dt}\Big/\frac{dy}{dx} = \mp\sqrt{T/\rho},$$

where the upper sign holds for waves propagated in the positive direction.

I have discussed this question at length, because it shows clearly the important fact that if waves are sent out from any disturbed region, the displacements in that region are not by themselves sufficient to determine the subsequent motion, the velocities being just as important as the displacements. In the above case, with the same displacements, the velocities might be chosen so as to give a wave wholly moving forward in one direction, or wholly moving back in the opposite direction, while generally there are two portions of the wave, one moving towards the positive, and one towards the negative side.

12. Transverse Waves in an Elastic Medium. We confine our attention for the present to bodies, the elastic properties of which are independent of direction. Such bodies are said to be "isotropic."*

Consider a medium in which the displacements are the same in magnitude and direction for all points lying in the same plane drawn normally to a given line. In Fig. 14 OX represents this line, and A_1B_1, A_2B_2, A_3B_3... are the intersections of a number of planes perpendicular to OX with the surface of the paper. At each point in these planes the displacements are supposed to be identical, but they may differ in different planes. If the displacements are all normal to OX and in the plane of the paper, each plane may be imagined to slide along itself through distances equal respectively to C_1C_1', C_2C_2' etc. We confine the investigation to the case of elastic forces which are such that for the linear displacements contemplated, the restitutional force acts backwards in the direction of the displacement.

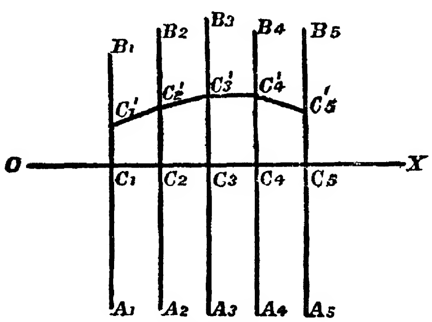


Fig. 14.

The strain set up in the medium by the displacement is one involving change of shape only, but not any change of volume.

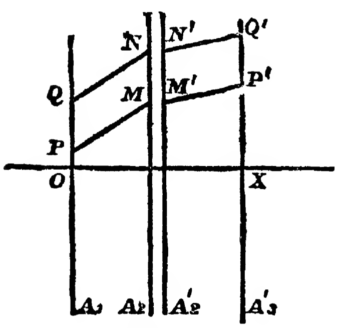


Fig. 15.

Let PM and $M'P'$ be the positions in the strained condition of two lines originally parallel to OX; the parallelogram $PQNM$ was originally a rectangle, and the elementary theory connecting strains and stresses shows that the plane A_2B_2 can only be maintained in its displaced position if it be acted on by an upward force which per unit surface is equal to $n \tan \alpha$, where n is the resistance to distortion and α the angle between PM and OX. Similarly the plane $A_2'B_2'$ to keep its position must be acted on by a downward force which per

* Thomson and Tait, Vol. I., Art. 676, give the following definition of isotropy:

"The substance of a homogeneous solid is called isotropic when a spherical portion of it, tested by any physical agency, exhibits no difference in quality however it is turned. Or, which amounts to the same, a cubical portion cut from any position in an isotropic body exhibits the same qualities relatively to each pair of parallel faces. Or two equal and similar portions cut from any positions in the body, not subject to the condition of parallelism, are undistinguishable from one another. A substance which is not isotropic, but exhibits differences of quality in different directions, is called eolotropic."

unit surface is $n \tan \alpha'$, where α' is the angle between $P'M'$ and OX. The constraint which maintains a small rectangular volume of unit height, thickness NN' and length MN in equilibrium is

$$-MN \times n(\tan \alpha' - \tan \alpha).$$

When the constraint is removed the volume will begin to move under the action of an elastic force equal and opposite to this.

If the displacements are denoted by y, we have

$$\tan \alpha = \frac{dy}{dx},$$

and if $MM' = t$,

$$\tan \alpha' = \frac{dy}{dx} + t \frac{d}{dx}\left(\frac{dy}{dx}\right),$$

so that the resultant elastic force may be written $MN \times t \times n \dfrac{d^2y}{dx^2}$, but $MN \times t$ is the volume considered, and if ρ is the density, $n \dfrac{d^2y}{dx^2}\Big/\rho$ will denote the resultant force divided by the mass which is equal to the acceleration. Therefore:

$$\frac{d^2y}{dt^2} = \frac{n}{\rho}\frac{d^2y}{dx^2}.$$

This equation is of the form (2) and shows that the medium is capable of transmitting waves in a direction OX with a velocity $\sqrt{n/\rho}$. As the velocity is independent of the wave-length, waves of any shape are propagated without change of type.

If we imagine a second disturbance superposed on the one which has been discussed, and at right angles to it, we obtain a wave propagation in which each particle describes a plane curve. We may for convenience limit the discussion to waves of the normal type, in which the displacements are therefore represented by

$$y = a \cos(\omega t - px).$$

Superposing a similar wave, with displacements in the z direction

$$z = b \cos(\omega t - px + \delta),$$

the paths of the particles in each plane are all similar and elliptic, circular or rectilinear, according to the value of δ and the relations holding between a and b (Art. 6).

One important observation remains to be made. Imagine the medium to consist of a number of detached particles, not acting on each other, but each attracted to its position of equilibrium by a force varying as the distance. Let the position and velocities of the particles at the time $t = 0$ be represented by

$$y = a \cos px$$

and

$$v = \omega a \sin px\,;$$

then the particles will continue to move under the action of the central forces so that their position at any subsequent time is represented by

$$y = a \cos(\omega t - px),$$

for this is the only relation which satisfies the condition that the accelerations are proportional to the displacements, and gives the required values for the displacements and velocities when $t = 0$. Hence a number of detached particles may simulate a wave motion, if once their displacements and velocities are properly adjusted, and if the force tending to bring them back to their position of rest causes an acceleration proportional to the displacement.

13. Condensational Waves. We imagine the same conditions to hold as in the previous paragraph, with the exception that the displacement (ξ) shall be in the direction of propagation. An investigation very similar to the one which was applied to the distortional or transverse waves will now hold, and it is not necessary to deduce again in detail the equation of motion, which for the case that $\frac{d\xi}{dx}$ is small is found to be:

$$\frac{d^2\xi}{dt^2} = \frac{m}{\rho}\frac{d^2\xi}{dx^2}.$$

Here m represents the longitudinal stress per unit elongation. It would be wrong to substitute for m the resistance to dilatation, or, as one might be tempted to do, Young's Modulus. The magnitude of m in terms of the elastic constants needs to be specially determined by the fact that there are no displacements at right angles to the direction of propagation. This we proceed to do. If the forces acting in the medium were all in the direction OX, a contraction of the medium at right angles to the direction of propagation would take place. The application of Young's Modulus would be justified in that case, but we have worked under the assumption that the displacements (*not* the forces) are parallel to OX. To counterbalance the contraction, transverse forces must act, and these forces will affect the elongations. It is known from the elementary theory that if P be the normal tensional stress along XO, it will produce an elongation equal to

$$P\left(\frac{1}{9k} + \frac{1}{3n}\right)$$

where k is the resistance to compression, and n the resistance to distortion.

The contraction at right angles is

$$P\left(\frac{1}{6n} - \frac{1}{9k}\right).$$

If equal tensions Q act along OY and OZ at right angles to OX, the elongations along OY and OZ are both equal to

$$Q\left(\frac{1}{9k}+\frac{1}{3n}\right)-(P+Q)\left(\frac{1}{6n}-\frac{1}{9k}\right)=Q\left(\frac{2}{9k}+\frac{1}{6n}\right)-P\left(\frac{1}{6n}-\frac{1}{9k}\right) \quad \text{..................(6).}$$

The elongation along OX is, taking account of Q,

$$P\left(\frac{1}{3n}+\frac{1}{9k}\right)-2Q\left(\frac{1}{6n}-\frac{1}{9k}\right).$$

Substituting the value of Q found by equating (6) to zero, the elongation becomes

$$\frac{3P}{3k+4n}.$$

The stress per unit elongation is therefore

$$m=k+\tfrac{4}{3}n.$$

The velocity of propagation is $\sqrt{m/\rho}$ and depends therefore on the resistance to distortion, as well as on the resistance to compression.

The waves which involve longitudinal displacements only, are called condensational waves, because they involve changes of volume, but all condensational waves involve also distortion. A difficulty may be found in admitting the existence of waves having the type considered on account of the force Q which would have to be applied at the boundary of the medium. The difficulty no doubt exists in some cases and it would be wrong, for instance, to apply the result obtained, to the propagation of waves along a rod or bar. Waves in which the displacements are solely in the direction of propagation could not travel along a rod, unless forces were applied at the surface and adjusted so as to prevent all contraction or expansion at right angles to the rod.

In an elastic medium, the boundaries of which are at a considerable distance, plane waves do not occur except as the limiting case of spherical waves, when the radius of the sphere has become very large. There is no difficulty in conceiving radial displacements and stresses across planes of AB and $A'B'$ (Fig. 16), which prevent the lateral contraction. Our investigation may therefore be considered to apply to such spherical waves having a large radius.

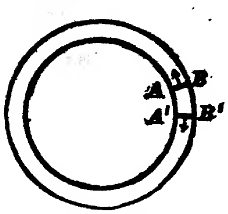


Fig. 16.

14. Spherical Waves. If a disturbance is produced within a small volume of an isotropic elastic medium, it spreads out in the form of spherical waves. Let at any one time, a very small volume T be disturbed, the rest of the medium being in a state of equilibrium. If all disturbing forces are now removed from the region T, the complete theory proves, what the results of the previous paragraphs already lead us to expect, that v being the velocity of propagation, the disturbance after a time t, will be confined to the neighbourhood of a spherical surface drawn with uniformly increasing radius $r = vt$ about some point within T. If the medium can propagate both distortional and condensational waves, the disturbance in general separates into two portions; one of these is spread over a sphere of radius $r_1 = v_1 t$, and consists of displacements which do not involve any change of volume, while the other, spread over the sphere of radius $r_2 = v_2 t$, involves both condensation and distortion. In terms of the elastic constants, the velocities of propagation are the same as for plane waves, so that

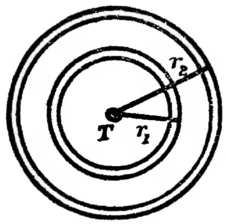


Fig. 17.

$$v_1 = \sqrt{n/\rho}, \qquad v_2 = \sqrt{(k + \tfrac{4}{3}n)/\rho} \quad \ldots\ldots\ldots\ldots\ldots\ldots (7).$$

In all fluid media, the resistance to change of shape is zero, hence the distortional wave does not exist, and the condensational wave is propagated with velocity $\sqrt{k/\rho}$, where for rapid oscillations, such as take place in sound waves, k is the adiabatic and not the isothermal elasticity. If a medium is incompressible, k is infinitely large, and the condensational wave is propagated with infinite velocity.

If the disturbance is of the normal periodic type, waves spread outward from the source, and, in consequence, energy is propagated outwards. Unless there is a continuous accumulation of energy in space, the energy passing in unit time through all closed surfaces surrounding the centre of disturbance, must be the same. Apply this to spheres of different radii drawn round the centre, when it will be clear that as the total energy transmitted through each sphere is the same, the energy per unit surface must be inversely proportional to the square of the distance.

Remembering (Art. 3) that the energy of a particle performing periodic oscillations is proportional to the square of the amplitude and following the analogy of plane waves, we are tempted to write for the displacements (y) in a spherical wave,

$$y = \frac{a}{r} \cos 2\pi \left(\frac{t}{\tau} - \frac{r}{\lambda} \right) \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (8),$$

where a is a constant which may be different for different directions, but remains the same along the same radius. This is not, however, the correct expression (Chapter XIII) though it is approximately accurate, when r is large compared to λ, and becomes more and more nearly true in proportion as $\lambda/2\pi r$ is negligible.

According to (8) the difference in phase between two points at a distance $r_2 - r_1$ from each other, along the same radius, would be $2\pi(r_2 - r_1)/\lambda$, but this result is limited by the same restrictions as the equation itself and need not be true unless r_1 is large compared with λ. Though the energy transmitted by a spherical wave must vary inversely as the square of the distance from the source, the energy of motion near the source is not necessarily all transmitted and hence the motion near the source is not correctly represented by a simple equation like (8).

15. Waves spreading from a disturbed region of finite size. If the original disturbance be spread over a space T of finite dimensions, Fig. 18, we may by a simple geometrical construction find the space which at any subsequent time t may be disturbed in consequence of the wave motion spreading out from T. We assume that no forces continue to act within T, that space being left to regain a state of equilibrium under the action of its own elastic forces only.

Subdivide T into indefinitely small portions and consider each small portion to be an independent centre of disturbance, from which spherical waves spread out as in the last paragraph. If A and B are the two points in T which are nearest and furthest, respectively, from P, then at a time AP/v the disturbance from A has just reached P. Previous to that time the point P was at rest. It will continue to be affected by waves coming from some point in T until a time equal to BP/v.

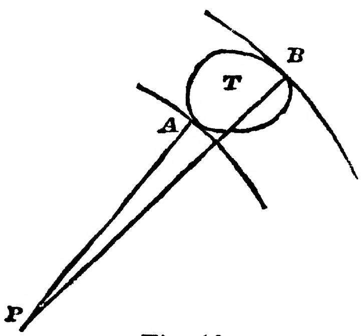


Fig. 18.

Then the disturbance will have completely passed over it, and P will again be in equilibrium, *i.e.* its velocity will remain zero, though its position may be different from that which it occupied previous to the passage of the wave. To obtain the region over which the disturbance is spread at any time t, we may draw spheres with radius vt, round every point of the boundary of T. These spheres will have one or two bounding envelopes, which separate the space cut by the spheres, from that which includes all points which are not cut by any sphere of radius vt drawn round any point within T as centre. The envelope or envelopes therefore form the boundary of the disturbed region. In Figures 19, 20 and 21 the disturbed space is supposed to have a rectangular section, and the sections of those waves are drawn which spread out from the edges of the disturbed region. In the first figure the time t is taken to be small, so that there is only one envelope and one boundary. In Fig. 20, t has increased sufficiently to show a space in the centre of the originally disturbed region, in which equilibrium has been restored. This space expands until as shown in Fig. 21 the disturbance is confined to a shell surrounding a considerable space in which the disturbance has ceased; the boundaries of the disturbed region approach the shape of spheres.

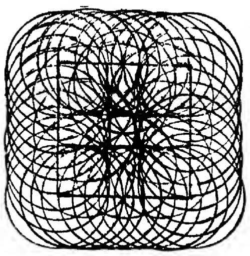


Fig. 19.

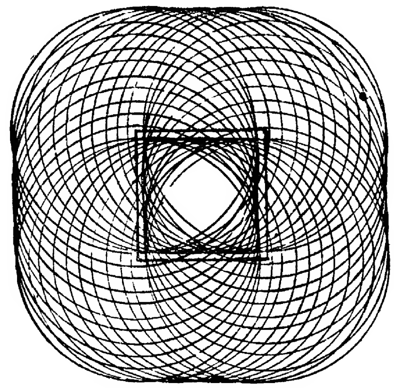


Fig. 20.

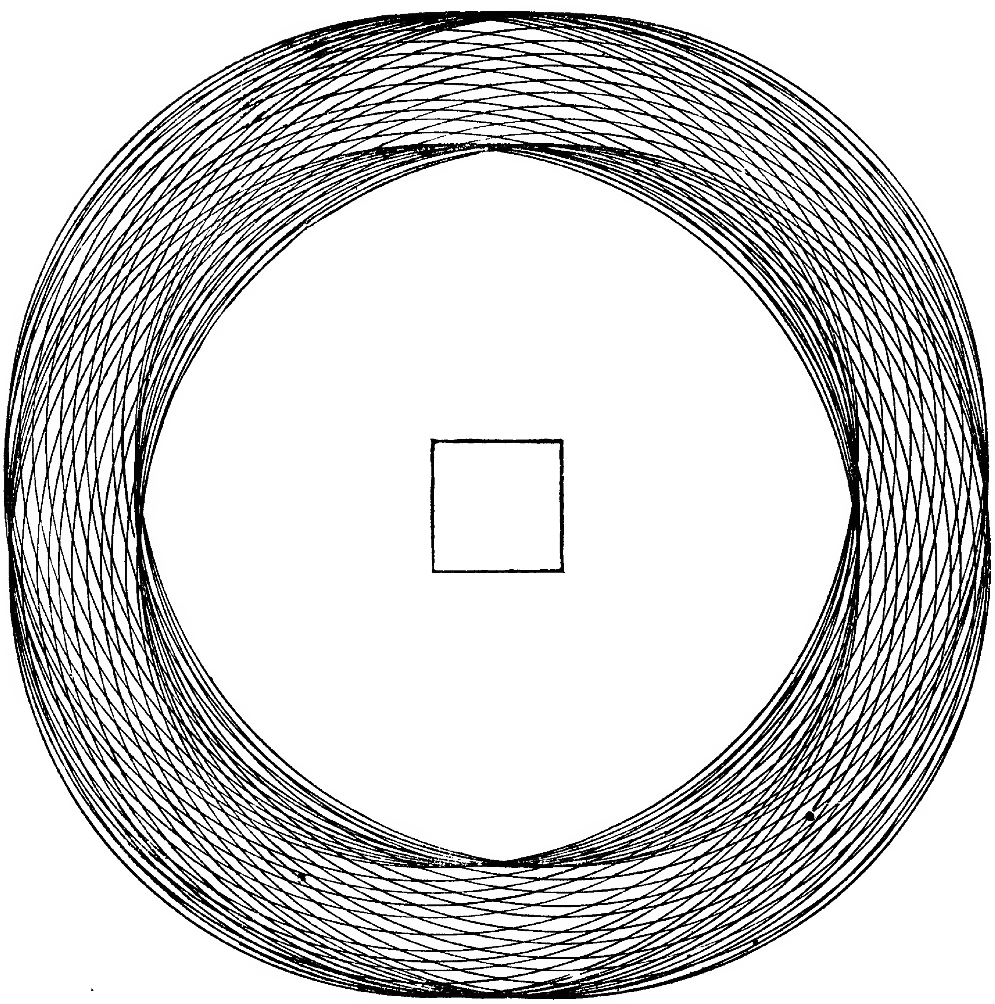


Fig. 21.

16. The Principle of Superposition. It has been assumed in the last article that the disturbance at P may be obtained by superposing the disturbances reaching it separately from all wave centres within T. This is called the principle of superposition, and holds, as may easily be proved, when the elastic properties of the medium are such that the stresses are linear functions of the displacements, or of their differential coefficients with respect to the coordinate axes.

In the special case discussed in Arts. 11 and 12, y being the displacement, the stresses are proportional to d^2y/dx^2, and satisfy therefore the condition of linearity. This still holds if the investigation is not limited to plane waves, as it was in these articles, for whatever be the properties of the medium, the stresses are always functions of the strains, and when the strains are small, their squares and products may ultimately be neglected. The principle of superposition may always therefore be taken to be an approximation which becomes more and more nearly true, the smaller the motion.

17. Huygens' Secondary Waves. Instead of following a disturbance from its original source, it is often more convenient to trace its subsequent propagation from its position and character at a given time. Thus let a disturbance originally coming from a small space be spread at time t over a thin spherical shell of which a portion AB is shown in Fig. 22. We may consider this shell to be the disturbed region and find the disturbance at time t_1 from Art. 15 by drawing spheres with radius $v(t_1-t)$ round each point of the shell. We get in this way two spherical envelopes $H'K'$ and HK between which the disturbance is necessarily confined.

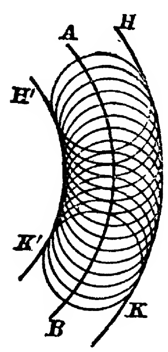


Fig. 22.

This result seems to be in contradiction with that obtained by another line of reasoning, for, going back to the original cause of the disturbance, the latter should, at time t_1, be confined to a thin shell of which HK is the outer boundary, and except close to HK there should be no disturbance.

This brings us to the important remark that the construction of Art. 15 only gives us the space in which there *may* be a disturbance, and not the space in which there is one necessarily. When the displacements and velocities of the originally disturbed regions are independent of each other, each point of the space in question will in general have a velocity and a displacement, and only in exceptional cases will these reduce to zero. But the displacements and velocities in the shell AB (Fig. 22) are not independent of each other, for they all originally came from the same source. Hence the waves which we may imagine to spread out from different points of AB must have some relation to each other as regards direction of displacement and velocity. As both our methods of reasoning are correct, it follows that the relation in the present instance must be such that there is neutralization at all points except in a narrow space close to the outer boundary HK.

If we imagine the velocities in AB to be reversed, the displacements remaining the same, we should get a wave travelling inwards. In that case, there should be neutralization of the secondary waves over HK and the disturbances would now lie in a shell close up to $H'K'$. This shows that the question whether a wave travels in one direction or another depends on the relation between velocities and displacements. The same result has already been proved in Art. 11.

The propagation of waves not necessarily plane or spherical may be treated in the same manner. So long as we know that the disturbance originally comes from a small space, and is therefore confined to a thin sheet, we may always obtain the disturbance at time t_1 from that at time t by constructing the outer envelope of all spheres having a radius $v(t_1 - t)$ and their centres on the boundary of the space to which the disturbance is confined at time t.

Huygens was the first to investigate the propagation of waves by means of secondary waves which he imagined to spread out from all points of the original wave, but the question why the disturbance should be confined to the outer envelope of the secondary spheres has been a serious difficulty up to the time of Fresnel, and even now the reason why, according to Huygens' construction, a wave should not be propagated backwards as well as forwards, is often a stumbling-block.

18. Refraction and Reflexion of waves. Imagine a plane wave disturbance to be confined to a narrow layer between two parallel planes of which AB and $A'B'$ are the intersections with the plane of the paper. Let this wave meet a surface HK which forms the boundary of another medium having similar properties to the first, but differing in the rate at which the waves travel through it.

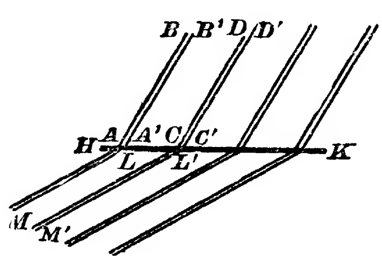


Fig. 23.

If in the second medium velocity of propagation were the same as in the first, the waves at time t_1 would be spread over a space between the parallel sheets, and it will now be shown that the wave on entering the second medium remains a plane wave, but with changed direction, so that LM, $L'M'$ may represent the boundaries of the space to which the wave has spread. To prove this, let AB (Fig. 24) represent the front of the sheet of disturbance which is supposed to be at right angles to the plane of the paper. After a time t the wave has moved forward in the first medium through a distance

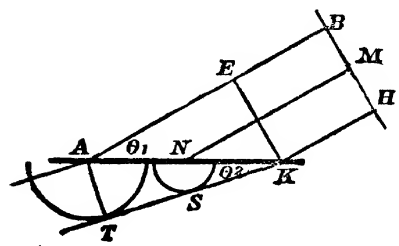


Fig. 24.

$$BH = v_1 t.$$

In the meantime, we may imagine, according to the previous articles, a secondary wave to have spread from A through a distance $v_2 t$, where v_2 is the velocity of propagation in the second medium. Draw therefore a sphere of radius $AT = v_2 t$. To trace another secondary wave we choose a time, say nt, at which the wave occupies in the first medium a position such that $BM = nBH$; its point of intersection with the line AK will be N, such that $AN = nAK$. From this point N, waves spread out, and at time t, *i.e.* an interval $t(1-n)$ after the wave has reached N, this secondary wave will have a radius $v_2 t(1-n)$. If all these secondary waves are drawn for values of n between 0 and 1, they are found to have a common tangent plane KST. This tangent plane gives the extreme limit of the disturbance in the second medium at the time t and represents therefore the wave-front at the time t. Draw KE normal to the wave in the first medium, AT normal to the wave in the second medium, and let θ_1 and θ_2 represent the angles between the wave and the surface of separation. Then an inspection of the figure gives

$$\frac{\sin\theta_1}{\sin\theta_2} = \frac{KE}{AT} = \frac{v_1}{v_2} \qquad \text{.........................(9).}$$

We call the wave in the second medium the refracted wave, and equation (9) gives the law of refraction. The "refractive index" of a substance as commonly defined is therefore equal to the ratio of the velocity of light in vacuo, to the velocity of light in the substance. The reflexion of waves may be treated exactly in the same manner, and the well-known law deduced, according to which incident and reflected waves are equally inclined to the surface of separation.

19. Wave Front and Wave Surface. In a medium in which waves of all periods are propagated with equal velocities, a wave-front is best defined as a surface such that the disturbance over it came originally from the same source, and started from that source at the same time. This does not restrict us to any particular form or shape of the wave. If the disturbance follows the law of normal oscillations the wave-fronts are also surfaces of equal phase. This follows from the fact that if we imagine ourselves to follow *e.g.* a condensation as it leaves a source, and spreads outwards with the velocity at which the wave is propagated, the locus of the condensation will, by the above definition, be a wave-front; it will also remain a locus of equal phase and remain so, though the wave may be refracted and reflected. When the medium transmits waves of different lengths with different velocities, the above definition no longer applies, and there is then strictly speaking no wave-front, though if we limit the discussion to homogeneous waves we may still speak of surfaces of equal phase, to which the term "wave-front" is sometimes conveniently though not very accurately applied.

A wave-front may lie altogether in one medium, or partly in one and partly in the other. Thus in Fig. 23 *DCLM* represents the trace of a wave-front. We apply the term wave *surface* to the front of a wave or to the surface of equal phase which completely surrounds a small centre of disturbance, but we confine the term to the case where the disturbance has never passed out of the original medium. A wave *surface* in a homogeneous medium like air, glass, or water, is always a sphere, while the shape of a wave-*front* would depend on the previous history of the wave, and might be plane, spherical or of irregular shape. In all media whether crystalline or isotropic the wave surface is characteristic of the medium, while the wave-front in general is not.

CHAPTER III.

PRELIMINARY DISCUSSION OF THE NATURE OF LIGHT AND ITS PROPAGATION.

20. The Nature of Light. For our present purpose we may consider light to be a wave-motion in an incompressible medium filling all space and permeating all bodies. We speak of the medium as the "luminiferous æther." The waves of light are distortional waves, the displacements being in the wave-front. Waves of the simple periodic form are propagated through the æther with a velocity independent of the wave-length. Hence any plane wave may be propagated without change of type.

A wave in which the displacements *at every point* are simply periodic, is called a homogeneous wave. If *e.g.* the displacement in a plane wave travelling in the direction of x is represented by

$$y = a \cos 2\pi \left(\frac{t}{\tau} - \frac{x}{\lambda}\right) \quad\text{.......................(1)},$$

without limitation as to the distance x, we should have a homogeneous vibration of wave-length λ, period τ, and frequency $1/\tau$. But we have no practical experience of a homogeneous wave of light. If it existed, *i.e.* if equation (1) were strictly true, the oscillation of any point would know no limit as regards time, either in the positive or negative direction. *A particle cannot send out homogeneous radiations unless it has been vibrating for an infinite time*, and the mere fact that we are lighting a flame, and extinguishing it, shows that the flame does not send out homogeneous radiations. Students should clearly realize that this is a consequence of our definition of homogeneous light. We cannot alter that definition without introducing a vagueness into our ideas, which has been the cause of much error and confusion.

Our perception of light depends on a physiological sensation, but the waves which are capable of producing this sensation are restricted to a definite range of frequency. There are radiations which have all the properties of luminous radiations, but which we cannot perceive by means of our eyes because their wave-length lies outside that range.

When we speak of the "spectrum" we include the whole range of radiation emitted by a radiating body, and we distinguish between the visible portion of the spectrum, which extends from the red to the violet, and the invisible portion which includes the wave-lengths which are too long to produce a visible sensation (infrared radiations) and those which are too short to produce a visible effect (ultraviolet radiations). A heated body emits radiations consisting of transverse waves, which when the temperature is low, belong entirely to the infrared portion of the spectrum. As the temperature increases, shorter waves are added to the radiation and increase in intensity both absolutely and relatively to the rays previously emitted. Ultimately the waves belonging to the visible portion of the spectrum begin to be included, when the body becomes red hot. A still further increase of temperature adds other visible and ultimately the ultraviolet radiations.

Table I. gives approximate values of the length of different waves.

TABLE I.

	cms.
Extreme Infrared radiation observed by Rubens and Aschkinass	·0061
Extreme Infrared in Solar Spectrum (Langley) ...	·00053
Longest waves photographed by Abney	·00020
Extreme Red of Visible portion	·000077
,, Violet ,, ,, ...	·000039
,, Ultraviolet transmitted through air ...	·000018
Extreme Ultraviolet observed in vacuo*	·000010

The electrical vibrations emitted by an electric spark are of the same nature as luminous radiations, but the shortest electrical wave we have been able to produce is four millimetres long†, *i.e.* about one hundred times longer than the longest observed infrared wave.

Though the homogeneous wave represented by (1) is a mathematical abstraction, the radiations of some gases, rendered luminous by electric discharges under reduced pressure, may for many purposes be considered homogeneous. Certain facts, however, which will be discussed in Art. 31, show that the homogeneity is far from being complete, and to prevent misunderstanding we shall call radiations of this kind: "quasi-homogeneous." If it be required to obtain a mathematical expression for quasi-homogeneous waves we may write (1) in the form

$$y_1 = a_1 \cos\{2\pi\lambda_1 (Vt - x) + \alpha_1\} \qquad (1a),$$

and form a number of similar equations varying slightly the values of a_1 and λ_1. The disturbance may then be expressed as the sum of

* Millikan, *Astrophysical Journal*, LII. p. 47 (1920).

† Lampa, *Wien. Ber. CV*, p. 1048 (1896).

the displacements given by the equations separately. The phase angle α may differ in successive equations by any value we please without affecting the question of homogeneity. If the variations of λ_1 cease to be small and wave-lengths are included which cover the whole range of observable radiations, we obtain white light. To secure continuity of the spectrum the values of the component wave-lengths λ_1, λ_2, etc. must be taken to be so near to each other that no spectroscope which can be constructed is able to resolve them.

We shall frequently use the above representation of quasi-homogeneous and of white light, but it is important to realise already at this stage that it is not the only one. Were light to consist of a succession of disturbances similar to the one represented in Fig. 10 we should still be able to represent it by means of homogeneous waves, with the help of Fourier's theorem (Art. 10). If the separate disturbances succeed each other irregularly we should get white light; the distribution of intensity over the different wave-lengths would depend on the shape of the disturbance, and would be the same for a single disturbance as for the whole series jointly. If therefore we chose the shape to be such as to give us the distribution we wish to represent, the present method of expressing white light is in every way identical with the previous one. In many cases the exact distribution of intensity does not matter and we may assume the whole duration of the disturbance to be indefinitely short: it then becomes an "impulse." The consideration of white light as a succession of impulses is very instructive and often simplifies calculations considerably, as we need only deal with a single impulse; while if we start from the homogeneous vibration we have to perform the summation for all wave-lengths before we can arrive at a final result. It must be noted that we are at present not concerned with the question how the light originates: we take the disturbance as it is and try to represent it analytically, and just as there are many ways of resolving a system of forces, so are there many ways of resolving the motion of light into elements with which we can deal analytically. The resolution by homogeneous waves is one, the resolution by impulses another. Whenever the two methods seem to yield different results, a mistake has been made in their application.

We have spoken of the "distribution of intensity" along the spectrum and a few words are necessary to explain what is meant by that term. We cannot assign intensity to an isolated wave-length but only to a collection of wave-lengths covering a finite range. Thus if $\delta\lambda$ represents a small range of wave-lengths we may say that the intensity of the light within that range is $A\delta\lambda$. A is a function of the wave-length and characterises the distribution of light through the spectrum. In theoretical investigations it is often convenient to consider a range of frequencies instead of a range of wave-lengths, and in that case writing $\lambda = V/\kappa$ we have

$$f(\lambda)\,\delta\lambda = V\kappa^{-2} f(V/\kappa)\,\delta\kappa.$$

The function defining the distribution of intensities is now $V\phi(\kappa)$, where

$$\phi(\kappa) = \kappa^{-2} f(V/\kappa).$$

Students should note that the form of the intensity curve depends on the independent variable which is chosen. They will then recognize that statements, which are often met with, such as that the maximum intensity of solar radiation is in the infrared or that the maximum luminous effect is in the yellow have no meaning unless it is specified whether you compare equal ranges of wave-lengths or equal ranges of frequencies. These statements are often based on observations in which a spectrum formed by a prism is used. But the dispersion of the prism compresses the red end as compared with the blue, and if a thermopile were moved along the spectrum we should compare together ranges for which neither $\delta\lambda$ nor $\delta\kappa$ but approximately $\delta\kappa^2$ is constant. When measurements on the intensity of light have to be made with prisms, the numbers should be reduced so as to give the distribution on some definite scale, and it should always be clearly stated whether the scale is one of equal wave-lengths or equal frequencies.

21. Velocity of Light. The experimental methods by means of which the velocity of light may be measured are described in elementary books. It will suffice here to record some of the results obtained. A good summary of the numbers found by different observers has been given by Michelson*, who combining his own figures derived from Foucault's method of the rotating mirror, with Newcomb's result, and Cornu's value obtained with Fizeau's toothed wheel, gives as the most probable number for the velocity of propagation in vacuo

$$299{,}890 \pm 60 \text{ km./sec.}$$

The experiments of Newcomb which probably form the most trustworthy single set of measurements gave

$$299{,}810 \pm 60 \text{ km./sec.}$$

If we weight the three sets of experiments according to their probable error we find in centimetres per second $2{\cdot}99859 \times 10^{10}$.

For many purposes it is sufficient to take the number 3×10^{10} for the velocity of light which throughout this book is denoted by V.

Michelson describes a combination of the two methods by means of which the accuracy might, in his opinion, be increased considerably. Should this be experimentally possible it will be necessary to take account of the fact that it is the group velocity in air and not the wave-velocity which is measured (see Art. 183).

* *Phil. Mag.* III. p. 330 (1902).

22. Intensity, illumination and energy of radiation. The "intensity" of radiation at a point in a pencil of rays is the energy conveyed per unit time and per unit surface through a small surface placed at right angles to the pencil. The measure of the energy is the heat generated when the whole radiation is absorbed.

The "illumination" at a point of a surface is the energy conveyed through the surface per unit time, and per unit surface.

There is a distinction between "intensity" and "illumination," because in the definition of the former the surface which receives the radiation is placed at right angles to the direction of propagation, while the term "illumination" can be applied to a surface placed obliquely. When dealing with a number of pencils which are inclined to each other we may still speak of the illumination at a surface which receives the pencil, but we cannot speak of the intensity of the combined pencils.

Intensity should always be measured in mechanical units, and includes all radiations visible or not, but when we speak of "illumination" we generally use the term in its literal sense confining it to that part of the radiation which produces a luminous effect in our eyes: arbitrary units are then more convenient in practice. It is occasionally convenient to apply the word "illumination" to radiation generally which includes both the infrared and the ultraviolet region; we should not hesitate to do so when this saves circumlocution or the introduction of another expression.

The "energy of radiation" in a volume is the excess of the energy which has entered the volume over that which has left it. It is not possible in general to measure the energy in a volume because the introduction of the measuring appliance would disturb the flow of energy, but in simple cases we may indirectly obtain a useful expression. Let a stream of radiant energy fall perpendicularly on one of the sides of a rectangular volume, having surface s and length l. Before the flow is constant—when as much energy leaves the volume as enters it—energy enters the first surface, and if the disturbance had not yet reached the opposite face an accumulation takes place within the volume. If E be the intensity of radiation, the excess of energy which enters the surface per unit time is Es, and as it takes a time l/V to cover the length of the volume the total accumulation of energy is Esl/V or $E\tau/V$ if τ be the volume. Hence the energy per unit volume is the intensity of radiation divided by the velocity of light. If a number of independent but constant streams of radiant energy traverse a volume we may similarly prove that the energy per unit volume at any point is equal to the total illumination over the complete surface of the volume divided by the velocity of light.

23. Optical length and optical distance. Two distances are said to be optically equivalent when light takes the same time to pass through them. The optical length of a path is defined to be the length of its equivalent in vacuo. If the path traverses several media, the total optical length is the sum of the optical lengths of all the different parts. Thus if v_1, v_2, v_3, etc. are the velocities of light and s_1, s_2, s_3, etc. the lengths of the paths in the various media, then the optical length is

$$V\left(\frac{s_1}{v_1}+\frac{s_2}{v_2}+\frac{s_3}{v_3}+\dots\right).$$

But by Art. 18, if μ_1, μ_2, μ_3 are the refractive indices,

$$\mu_1 = V/v_1\,;\ \mu_2 = V/v_2\,;\ \mu_3 = V/v_3\,;$$

and hence the optical length of the path is

$$\mu_1 s_1 + \mu_2 s_2 + \mu_3 s_3 + \dots \qquad \dots\dots\dots\dots\dots\dots\dots(2).$$

The optical *distance* between two points is defined to be the shortest optical length of any curved, straight, or broken path that can be drawn between them. If both points lie in the same medium, the shortest path is clearly the straight line which joins them, and the optical distance is the length of this line multiplied by the refractive index of the substance.

A "ray" is defined to be a path of shortest optical length. In a medium possessing uniform optical properties, a ray passing through two given points, must, by this definition, always be the straight line which joins them. The path of a ray between two points which are situated in different media may be determined as follows:

Let A and R, Fig. 25, be the two points, and S some point on the surface of separation, which lies in the plane drawn through A and R, perpendicular to the surface. Draw AC perpendicular to AS, and RE perpendicular to SR. From any point T in the plane ASR draw TC parallel to AS, and TE parallel to SR, and construct perpendiculars SH and KT from S and T, on CT and SR respectively. Let the position of S be such that the optical length HT is equal to the optical length SK, then the optical length of $CT+TE$ is equal to that of $AS+SR$. But from inspection of the figure, $AT>CT$, $RT>TE$, hence the optical length of the path $AT+TR$ must be greater than that of the path $CT+TE$ which is equal to that of $AS+SR$. Students should convince themselves that the same result follows when the point T is taken to lie on the other side of S. It follows that the optical length $AS+SR$ is smaller than that of any other path joining A and R in the plane of the paper. The condition on which this result depends is that the

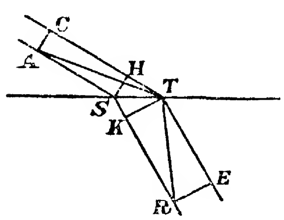


Fig. 25.

optical length of HT is equal to that of SK or that if μ_1, μ_2 are the two refractive indices,

$$\mu_1 HT = \mu_2 SK.$$

If θ_1 and θ_2 are the angles which AS and SR form with the normal to the surface, the condition reduces to

$$\mu_1 \sin \theta_1 = \mu_2 \sin \theta_2,$$

which is the well-known law of refraction. The rays as defined by us are therefore identical with the rays of geometrical optics.

It has been assumed in the above proof, that the path of shortest optical distance lies in the plane which is at right angles to the surface separating the two media. The restriction may be removed by giving to S a small displacement to either side at right angles to that plane, and showing that the optical distances AS and SR are both increased.

A ray may be drawn between any two points of an optical system, but only a single set of rays belong to one set of wave-fronts. Let HK and $H'K'$ (Fig. 26) represent two wave-fronts of the same disturbance. From a point A on HK, a line may be drawn tracing the shortest optical length between A and any given point C on $H'K'$. By altering the position of C, its optical distance from A changes, and some point may be found on $H'K'$ for which that optical distance is least. Let B be that point. The path of shortest optical length between A and B is one ray of the system which belongs to the two wave-fronts. We may similarly trace a ray satisfying the same conditions from every point P on HK to a corresponding point Q on $H'K'$, and thus obtain the system of rays belonging to a given system of wave-fronts.

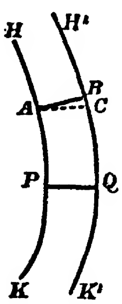


Fig. 26.

If the medium is homogeneous, the rays must be straight lines. In a number of separate media, each being homogeneous, the system of rays is made up of a system of straight lines, which will in general change in direction when passing from one medium to another.

If the medium is isotropic, so that one wave-front may be obtained from another by Huygens' construction, as explained in Art. 17, the system of rays intersects the system of wave-fronts at right angles. This is proved by considering two points, A_1, A_2, on a wave-front HK. Every other wave-front $H'K'$ will be a tangent surface to two spheres, drawn with the same radius round A_1 and A_2 as centres, so that if B_1, B_2, are the two points of contact, A_1B_1 and A_2B_2 must be at right angles to $H'K'$. This being so, A_2B_1 is necessarily longer than A_1B_1, provided that A_2 is sufficiently near to A_1. Hence all points on HK which are near

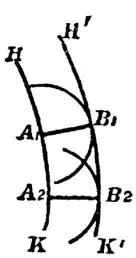


Fig. 27.

A_1 are further from B_1 than A_1, and therefore the sphere which is drawn through A_1 round B_1 as centre, cannot intersect, but must touch the surface HK. A_1B_1 stands therefore at right angles both to HK and to $H'K'$.

If the medium is isotropic, but not homogeneous, as *e.g.* the air surrounding the earth, which varies in density and temperature, the course of a ray may be curved, but the above proof still holds if we take HK, $H'K'$ to lie near each other, and hence the rays are in this case also at right angles to the wave-fronts.

It also follows from Huygens' construction that the optical length from one wave-front to another is the same when measured along different rays. We shall call this length the optical distance between the two wave-fronts.

To illustrate the use which may be made of these propositions, we may deduce the well-known formula connecting the position of a small object with that of its image formed by a lens.

If waves spread out from a point source at P, the wave-fronts are spheres with the point as centre. If these wave-fronts, after passing through the lens, are spheres with Q as centre, the wave-fronts will gradually contract until the energy of the waves is concentrated at Q. (This is not quite correct, owing to the fact that the wave-fronts after emergence are not *complete* spheres, but this does not affect the argument.) The optical length from P to any point on HK is the same, and also the optical length from any point on $H'K'$ to Q. It has been proved above that the optical distance from HK to $H'K'$ is the same when measured along any ray, hence the optical distance from an object to its image is the same along all rays. PSQ and $PMNQ$ are clearly lines satisfying all conditions laid down for the rays belonging to the system. If μ is the refractive index of the lens, the equality of optical lengths leads to the equation

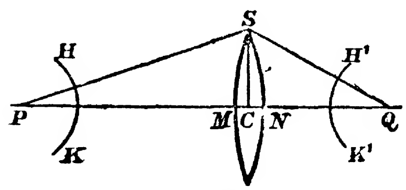


Fig. 28.

$$PS + SQ = PM + \mu MN + NQ = PQ + (\mu - 1)\, MN,$$

or

$$(PS - PC) + (SQ - QC) = (\mu - 1)\, MN.$$

Also

$$PS^2 - PC^2 = SC^2,$$

$$\therefore\ PS - PC = \frac{SC^2}{PS + PC} = \frac{SC^2}{2PC},$$

if the angle SPC is so small that its square may be neglected.

Similarly

$$SQ - QC = \frac{SC^2}{2CQ},$$

$$\therefore\ \frac{1}{PC} + \frac{1}{CQ} = \frac{2(\mu - 1)MN}{SC^2}.$$

If MN and SC are expressed in terms of the radii of curvature of the surfaces of the lens, we obtain the well-known relation between the position of object and image.

24. Fermat's Principle and its application. Fermat (1608 to 1668) enunciated the principle that Nature could not be wasteful, and was bound for this reason to cause the rays of light to travel between two points in the shortest time possible. We may accept Fermat's conclusions but cannot attach any weight to his reasoning, more especially as the mathematical conditions which are made use of in the application of his principle would hold equally if the optical distance between two points were a maximum instead of a minimum: it is indeed sufficient to define the ray as the line satisfying the condition that the optical length between two points is *stationary.*

"*Fermat's Principle,*" as it is called, often furnishes a powerful method of dealing quickly with otherwise complicated problems, and may serve as a connecting link between the waves of the undulatory theory and the rays of Geometrical Optics. The importance of the property of stationary optical length lies in the fact that it enables us often to determine the optical distances with sufficient accuracy when the course of the ray is only approximately known. That the optical length is the same when measured along a ray or a line infinitely near the ray, follows directly from the maximum-minimum property.

HK, $H'K'$ being wave-fronts, let APB be a ray belonging to the system. Let AQB be a line lying everywhere near APB so that their distance apart ST at any point S may be expressed in terms of the position of S, and the separation PQ at some definite point Q. Writing $PQ = \alpha$, the difference in the optical length of AQB and APB must then be expressible in terms of α, and if α is small, must be capable of expansion in a series proceeding by powers of α as *e.g.*

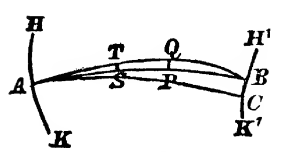


Fig. 29.

$$h_1\alpha + h_2\alpha^2 + h_3\alpha^3 + \ldots\ldots$$

As the condition of stationary optical path requires that for numerically equal positive and negative small values of α the optical length shall be the same and place h must be zero. If the coefficients of the higher powers of α also vanish, B will be the image of A, and if for each point of HK we can find a point on $H'K'$ fulfilling the same conditions, the surface $H'K'$ is called "aplanatic" with respect to HK.

It is important to notice that the stationary property also holds for two paths AB and AC, if B and C are points near each other and on the same wave-front.

For this purpose it is only necessary to consider that if the difference between the optical lengths of AB and AC depended on the first power of BC, the reversal of the direction of BC would change the sign of the difference. Hence a position of C could be found for which the optical length AC is shorter than that of the ray AB; but this would not be compatible with the construction of the wave-front $H'K'$ from the wave-front HK according to Huyghens' principle (Art. 23).

The application of Fermat's principle may be illustrated by an example.

Let a parallel beam (*i.e.* a beam in which the rays are parallel and therefore the wave-fronts planes which cut the rays at right angles) fall on a prism and be refracted through it. Let HK and LM (Fig. 30) be two wave-fronts, then μ being the refractive index, the equality of optical lengths gives

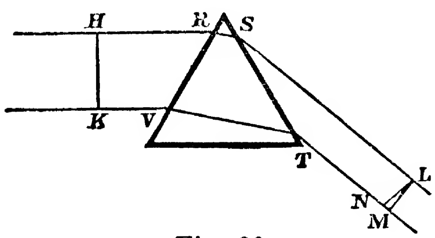


Fig. 30.

$$HR + \mu RS + SL = KV + \mu VT + TM.$$

Suppose a wave of slightly different wave-length and refractive index μ' falls on the prism, the incident beam being coincident with that just considered. We may take HK to be also one of the fronts of the second set of waves, but on emergence, the wave-fronts of the wave defined by μ' will not be parallel to those defined by μ. We select that front which passes through L. Let its inclination be such that it intersects the ray TM in N. If μ' and μ only differ by a small quantity, we may measure the optical length of any of the rays μ' not along its own path, which we do not know, but along the path traced out by one of the rays μ which lies near it, the error committed depending only on the second power of $\mu' - \mu$. We may therefore obtain a second equation for the equality of optical lengths, which is

$$HR + \mu' RS + SL = KV + \mu' VT + TN;$$

taking the difference between the two equations,

$$(\mu' - \mu) RS = (\mu' - \mu) VT - MN,$$

or

$$(\mu' - \mu)(VT - RS) = MN.$$

The angle θ formed between the emergent rays of the two beams is equal to the angle NLM, or if small, equal to its tangent NM/ML. It follows that

$$\frac{\theta}{(\mu' - \mu)} = \frac{VT - RS}{ML} \quad \text{.......................(3).}$$

This is a useful expression, first obtained by Lord Rayleigh, connecting the dispersion of a prism with the width of the emergent beam, and the lengths of the paths traversed in the prism by the extreme rays of the beam.

25. The Principle of Reversibility. There is an important proposition affirming that if at any time all velocities in a dynamical system are reversed and there is no dissipation of energy, the whole previous motion is reversed. Any configuration of the system which existed at a time t before the reversal took place will therefore again exist at the time t after reversal.

As an example of this principle, I give an investigation originally due to Stokes, which yields important relations between the amplitudes of incident, reflected and refracted light. Let a ray of homogeneous light AO (Fig. 31), of unit amplitude, fall on a reflecting surface. Let r be the amplitude of the reflected ray OR, and t that of the transmitted ray OT. If at any moment the courses of the reflected and refracted rays are reversed, the two reversed rays coming together at the surface should combine to reproduce the ray of unit amplitude passing along OA and nothing else. That is to say, the ray OT' due to the reflexion of TO must be neutralized by the ray due to the refraction of RO. We shall begin by assuming that there is no change of phase at reflexion or refraction, except possibly one of 180° which will appear as a reversal of the sign of the amplitude. If r' measures the amplitude after reflexion at O of a ray of unit amplitude travelling along TO, the ray which originally travelled along AO with unit amplitude, and after refraction took an amplitude t, will, after reversal and reflexion at O, have an amplitude tr'. Similarly the ray OR reversed and refracted takes an amplitude rt. Hence one of the conclusions we may draw from the principle of reversion is that

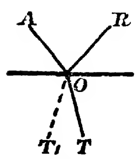


Fig. 31.

$$rt + r't = 0$$

or

$$r + r' = 0 \quad\text{..............................(4).}$$

This equation must be interpreted to mean that there is a reversal of phase either at internal or external reflexion, r' being equal in magnitude to r, but of opposite sign.

The ray OR of amplitude r, has after reversal and renewed reflexion at O, an amplitude r^2, the ray OT of amplitude t has, after reversal and refraction at O, an amplitude tt', if t' is the ratio in amplitude of the incident and refracted ray when the ray passes through the surface in the reverse direction. If the two rays make up the original one of unit amplitude, it follows that

$$r^2 + tt' = 1 \quad\text{..............................(5).}$$

The equations are not sufficient to determine t and t' in terms of r, but they establish an important relation.

We may now generalize our results so as to include the possibility of a change of phase.

Let the oscillation in the incident ray at the point of contact with the reflecting surface be given by the projection on a fixed line of the revolution of a point I (Fig. 32) in a circle, the revolution being counter-clockwise. Let similarly the motion at the same point of the reflected and refracted waves be represented by the projection of the circular motion of R and T. The system of points I, R, T revolving with the same angular velocity represents at any time, the phases at the point of incidence of the incident, reflected and refracted rays. At some instant let the rays be reversed; the effect on our diagram will be that the points T and R now revolve clockwise, but their position at the time of reversal is unchanged. The reflected ray reversed will give a reflected ray represented by OR_1 where R_1 must lie on OI, because obviously the reflexion in the reverse direction must produce a change of phase which is identical in magnitude with the change of phase in the forward direction. The refracted ray OT gives rise, on reversal, to a refracted ray, which again must be capable of representation as a projection of the circular motion of some point T_1 and this point must also lie on OI because the principle of reversion shows that OR_1 and OT_1 must have the resultant OI.

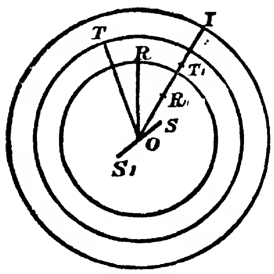


Fig. 32.

With the same notation as before, we find that the equation

$$r^2 + tt' = 1$$

is independent of any assumption as to change of phase at reflexion.

The reflected wave OR gives rise after its reversal to a refracted wave which as regards phase and amplitude, may be represented by OS where $OS = tr$, while the refracted wave OT gives rise to a reflected wave represented by the vector $OS_1 = tr'$, which must neutralize OS. This as regards magnitude leads to the equation

$$r = r'.$$

Now the angle ROS must be equal to IOT, and

$$\begin{aligned} TOS &= TOI + IOS \\ &= TOI + ROS - ROI \\ &= 2TOI - ROI. \end{aligned}$$

If the change of phase (IOT) at transmission be denoted by τ, and the change of phase at reflexion (IOR) by ρ, then the change of phase ρ' at the internal reflexion is TOS_1, measured clockwise, which is the direction of the reversed motion; this is equal to $\pi + TOS$.

Hence $$\rho' = \pi + 2\tau - \rho$$

or $$\rho + \rho' = \pi + 2\tau$$

gives the complete law, which for $\tau = 0$ reduces to $\rho + \rho' = \pi$ as previously established. The change of phase at transmission is by the same reasoning shown to be the same in whichever direction the refraction takes place. This completes the information we can get out of the principle of reversibility in dealing with this problem.

26. Polarization. Equation (1) represents a disturbance in which the displacement is everywhere in the same direction. A ray of light satisfying this condition is said to be plane polarized. If the path of the displaced point is circular or elliptical we say that the light is circularly or elliptically polarized. It is observed that when a ray of light AR (Fig. 33) is reflected from a glass surface HK, at a particular angle, and the reflected ray RS is incident on a second mirror LM, which is capable of rotation round RS, the intensity of the reflected ray SB depends on the position of the second mirror. If LM be parallel to HK, the intensity of the reflected ray is a maximum, and if the mirror be turned through a right angle, so that the plane of incidence, instead of being in the plane of the paper, is at right angles to it, the intensity of the reflected light is zero. Such a result has no analogy in sound and could not be explained if light were due to longitudinal waves. In the case of transverse disturbances, we may draw a distinction between the vibrations which lie in the plane of incidence and those at right angles to it, and thus explain the want of symmetry. If we imagine that at a particular incidence, those rays only are reflected in which the vibration is at right angles to the plane of incidence, the ray RS will consist of vibrations at right angles to this plane and will be reflected in the same proportion by LM, if the two mirrors are parallel. But if LM be turned through a right angle, the vibration along RS will now be *in* the plane of incidence of the second mirror, and hence by hypothesis, no light is reflected. Light which has been polarized by reflexion, is said to be polarized *in* the plane of incidence.

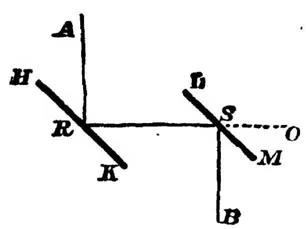


Fig. 33.

All homogeneous rays are polarized. To prove this, we imagine the wave to proceed in the direction of the axis of x.

Let the displacement have one component along OY,

$$y = a \cos \omega t,$$

and one along OZ,

$$z = b \cos \omega (t - \theta).$$

According to Art. 6 the motion is rectilinear when $\theta = 0$, circular when $a = b$ and $\theta = \pm \pi/2$, and elliptical in all other cases. Homogeneous light may therefore be plane, circularly, or elliptically polarized, but it will always be polarized.

We have certain experimental methods of detecting polarization. When these methods are applied to light emitted from a flame or from a body rendered incandescent by the electric discharge, it is found that no polarization can be detected, even when the lines of the spectra of gases are examined. This alone is sufficient to show, as has already been pointed out, that we do not meet with homogeneous light in nature.

If the path of a point over which a homogeneous wave passes be circular or elliptical we say the light is circularly or elliptically polarized. The definition of plane polarization can be applied to white light, if the disturbance is everywhere in the same direction, but the definition of circular and elliptic polarization as given above only applies to the ideal case of homogeneous light. With quasi-homogeneous light the path may be described as an ellipse gradually changing in dimension but preserving the same ratio of semi-axes and also the same direction of these axes. The changes of amplitude must be considered as slow compared with the time of one revolution, but rapid compared with the time in which we can carry out an observation, so that our instruments only perceive a general average effect. We say that white light shows circular or elliptical polarization when its quasi-homogeneous constituents have these properties.

27. Light reflected from transparent substances. It will be useful to follow out a little more closely at this stage the effects of reflexion from a transparent polished surface. According to the preceding article, ordinary light reflected by such a surface at a particular angle, called the angle of polarization, is plane polarized and by definition, polarized in the plane of incidence. Anticipating the results of later Chapters we specify at once, that the direction of vibration is at right angles to what has been called the plane of polarization.

The amplitude of the reflected light must, according to what has been said, depend (1) on the direction of polarization of the incident light, and (2) on the angle of incidence. A mathematical expression for the reflected amplitudes in different cases was first obtained by Fresnel, whose results we introduce here, deferring their theoretical demonstration to a later stage. If a homogeneous vibration of unit amplitude vibrating normally to the plane of incidence falls on a reflecting transparent substance, the angle of incidence and refraction being θ and θ_1 respectively, the amplitude of the reflected ray is:

$$r_n = \frac{\sin(\theta_1 - \theta)}{\sin(\theta_1 + \theta)} \qquad \text{(6)}.$$

If the light vibrates parallel to the plane of incidence the reflected vibration has an amplitude:

$$\pm r_p = \frac{\tan(\theta_1 - \theta)}{\tan(\theta_1 + \theta)} \quad \text{.......................(7).}$$

In these equations θ_1 denotes the angle of refraction so that if μ is the refractive index, $\sin\theta = \mu \sin\theta_1$. For the present we take these equations to represent experimental facts and apply them to particular cases. The sign of the amplitude is left undetermined for the present because we can only observe intensities which depend on the square of the amplitude. An experimental method to decide between the alternatives is given in Art. 44.

The square of r_n increases with increasing incidence from $\theta = 0$ (normal incidence) to $\theta = \pi/2$ (grazing incidence). When θ is sufficiently small, we may put $\sin\theta = \theta$, $\theta = \mu\theta_1$, and obtain

$$r_n = \frac{1 - \mu}{1 + \mu} \quad \text{.............................(8).}$$

This holds for normal incidence and gives us the intensity of the reflected light at that incidence:

$$\left(\frac{\mu - 1}{\mu + 1}\right)^2.$$

Thus for glass with refractive index 1·5, one twenty-fifth or 4% is reflected at normal incidence, and hence 96% is transmitted. When the incident ray is as oblique as possible, the light is entirely reflected, none being transmitted. The negative sign of r_n when μ is greater than one indicates a change of phase of 180°. The expression for the amplitude of the light polarized at right angles to the plane of incidence diminishes from

$$r_p = (1 - \mu)/(1 + \mu),$$

for normal incidence, to 0, when $\theta + \theta_1 = \pi/2$. In that case $\sin\theta_1 = \cos\theta$, and the equation of refraction $\sin\theta = \mu\sin\theta_1$ becomes $\tan\theta = \mu$. If the angle of incidence further increases, the amplitude increases again and for grazing incidence the light is in this case also totally reflected.

Equations (6) and (7) preserve their numerical value, but reverse their sign when θ and θ_1 are interchanged. This shows that on reversal of the ray the same fraction of light is reflected, but that if in one case there is no change of phase, a change of 180° takes place in the other case. This agrees with the result independently deduced in Art. 25.

If the incident light has an amplitude a and is polarized in a plane inclined at an angle α to the plane of incidence, we may decompose the oscillations into two, one $a\cos\alpha$ being polarized in the plane of incidence and the other $a\sin\alpha$ polarized at right angles to that plane. The reflected rays of each component may then be united again. If b be the amplitude of the reflected ray, and β the angle its plane of polarization forms with the plane of incidence, we have

$$b \cos \beta = a \cos \alpha \frac{\sin(\theta_1 - \theta)}{\sin(\theta_1 + \theta)},$$

$$b \sin \beta = a \sin \alpha \frac{\tan(\theta_1 - \theta)}{\tan(\theta_1 + \theta)}.$$

Hence
$$\tan \beta = \tan \alpha \frac{\cos(\theta_1 + \theta)}{\cos(\theta_1 - \theta)},$$

and
$$b^2 = a^2 \left\{ \cos^2 \alpha \frac{\sin^2(\theta_1 - \theta)}{\sin^2(\theta_1 + \theta)} + \sin^2 \alpha \frac{\tan^2(\theta_1 - \theta)}{\tan^2(\theta_1 + \theta)} \right\}.$$

The first of these equations shows that for $\theta + \theta_1 = \pi/2$, the reflected ray is polarized entirely in the plane of incidence. The resulting value of θ obtained from $\tan \theta = \mu$, gives us therefore the angle of polarization.

When ordinary light falls on a reflecting surface, we may obtain the intensity of the reflected light by considering that the homogeneous waves of closely adjoining wave-lengths have their planes of polarization distributed quite irregularly; $\cos^2 \alpha$ and $\sin^2 \alpha$ in the above equation must therefore be replaced by their average value, which is one half. The intensity of the reflected light is therefore

$$b^2 = \tfrac{1}{2} a^2 \frac{\sin^2(\theta_1 - \theta)}{\sin^2(\theta_1 + \theta)} \left(1 + \frac{\cos^2(\theta_1 + \theta)}{\cos^2(\theta_1 - \theta)} \right).$$

The total intensity is given by this expression, but the intensity is distributed unsymmetrically in different directions. That part of the light which is polarized at right angles to the plane of incidence has an intensity

$$\tfrac{1}{2} a^2 \frac{\tan^2(\theta_1 - \theta)}{\tan^2(\theta_1 + \theta)},$$

while for the light polarized in the plane of incidence, the intensity is

$$\tfrac{1}{2} a^2 \frac{\sin^2(\theta_1 - \theta)}{\sin^2(\theta_1 + \theta)}.$$

The difference between these two quantities gives us the amount of polarized light, which, together with the unpolarized light of intensity equal to twice that of the smaller, makes up the partially polarized beam of the reflected light.

The intensities of the transmitted beams are obtained by the principle of the conservation of energy, and if I_a, I_r, I_t represent the intensities of the incident, reflected, and transmitted beams respectively,

$$I_a = I_r + I_t.$$

It would be wrong to conclude from this that if a_a, a_r, a_t measure the *amplitudes* of the incident, reflected, and transmitted rays, $a_a^2 = a_r^2 + a_t^2$, because the squares of amplitudes only express the relative intensities if the waves have the same wave-length, and are transmitted through media possessing the same inertia. It is, however, in every case the intensity that concerns us, and the equations given above give therefore everything that is required.

So far as can be judged by experiment, Fresnel's equations (6) and (7) represent the observed facts with considerable accuracy. An important exception is that in which the angle of incidence lies near the angle of polarization. If the incident light be polarized at right angles to the plane of incidence, and falls on the surface at the angle of polarization, no light should according to equation (7) be reflected at all, and there should be a complete reversal of phase in the reflected light, as the angle of incidence changes from a value slightly smaller than the angle of polarization to a value slightly greater. Sir George Airy discovered that this is not quite correct for highly refracting substances like diamond, and Jamin, pursuing the subject further, found that there is always a residue of light reflected at the polarizing angle though the incident light may be strictly polarized at right angles to the plane of incidence. The phase, which should change suddenly through 180°, changes rapidly but not discontinuously, so that at the polarizing angle there is a retardation or acceleration of phase amounting to 90°.

Since then, Lord Rayleigh* has shown that Jamin's results are in great part, though not entirely, due to surface films of probably greasy matter which may be removed by polishing.

If light falls on the surface of a plate of glass at the polarizing angle, the ray entering the glass falls on the second surface again at the polarizing angle, as the condition $\theta + \theta_1 = \pi/2$ will, in a plate bounded by parallel surfaces, be fulfilled at both incidences. It follows that the light, reflected at the second surface, increases the intensity without detracting from the polarization of the reflected beam. The same argument may be used to show that a pile of parallel plates gives at the proper angle a polarized reflected beam which, neglecting absorption, might be made to equal the intensity of that component of the incident beam which is polarized in the plane of incidence. Such a pile furnishes a simple and cheap method of obtaining polarized light. There is some disadvantage, however, in the fact that the direction of the rays is changed by reflexion. For this reason, the transmitted beam is occasionally used. The transmitted beam is only partially polarized by a single refraction, but it is clear that when the number of plates is sufficiently great to reflect all the light polarized in the

* *Collected Works*, Vol. II. p. 522.

plane of incidence, the refracted beam can only contain light polarized at right angles to that plane. A large number of plates is however required, if the polarization is to be approximately complete. The amount of light transmitted through a pile of plates, or reflected from it, has been calculated by Provostaye and Desains*.

If ρ be the fraction of the intensity reflected at one surface, that reflected from a number n of parallel surfaces is

$$\frac{n\rho}{1+(n-1)\rho}.$$

If there are m plates, there are $2m$ surfaces, hence in terms of m the intensity of the reflected light is

$$\frac{2m\rho}{1+(2m-1)\rho}$$

and the intensity of the transmitted light is

$$\frac{1-\rho}{1+(2m-1)\rho}.$$

For glass of refractive index 1·54, ρ at the polarizing angle is ·16, and from this we may calculate that it requires 24 plates to furnish a transmitted beam which shall contain not more than 10 % of unpolarized light.

28. Total reflexion. When a ray is totally reflected, there is no refracted ray, but equations (6) and (7) still hold, provided we give to the angle of refraction the imaginary value which it takes according to the laws of refraction, interpreting amplitude, when it contains an imaginary term, according to principles explained in Art. 8. If θ denote the angle of incidence, in a medium of refractive index μ, the second medium being air, the law of refraction is

$$\mu \sin\theta = \sin\theta_1$$

and total reflexion takes place if $\sin\theta > 1/\mu$. In that case we may separate the imaginary and real parts for light vibrating normally to the plane of incidence as follows:

$$\sin(\theta_1-\theta) = \sin\theta_1\cos\theta - \cos\theta_1\sin\theta,$$

$$\sin(\theta_1+\theta) = \sin\theta_1\cos\theta + \cos\theta_1\sin\theta;$$

$$\therefore\ r_n = \frac{\sin\theta_1\cos\theta - \cos\theta_1\sin\theta}{\sin\theta_1\cos\theta + \cos\theta_1\sin\theta}.$$

All quantities are real except $\cos\theta_1$. The expression for r_n is of the form $(p-iq)/(p+iq)$ and hence, according to Art. 8, the amplitude is one. This was to be expected, since we are dealing with total reflexion

* *Ann. de Chemie et Phys.* xxx. p. 159 (1850).

Under these conditions, the complex amplitude is of the form $e^{i\delta_1}$, and its real part measures the cosine of the change of phase (δ_1). The real part of $(p-iq)/(p+iq)$ being $(p^2-q^2)/(p^2+q^2)$ we find, by reference to Art. 8,

$$\cos\delta_1 = \frac{1+\mu^2-2\mu^2\sin^2\theta}{\mu^2-1} \qquad \text{.................(9)}.$$

As special cases we have

$$\text{for } \sin\theta = \frac{1}{\mu};\ \delta_1 = 0,$$

$$\text{for } \theta = \frac{\pi}{2};\ \delta_1 = \pi.$$

This shows that at incipient total reflexion there is no change of phase and at grazing incidence, a reversal of phase.

In order to reduce the tangent formula, we transform as follows :

$$\frac{\tan(\theta_1-\theta)}{\tan(\theta_1+\theta)} = \frac{\sin 2\theta_1 - \sin 2\theta}{\sin 2\theta_1 + \sin 2\theta}$$

$$= \frac{(\sin 2\theta_1 - \sin 2\theta)^2}{\sin^2 2\theta_1 - \sin^2 2\theta}.$$

Here $\sin 2\theta_1$ is imaginary, but its square is real, hence for the real portion of the fraction we have

$$\cos\delta_2 = \frac{\sin^2 2\theta_1 + \sin^2 2\theta}{\sin^2 2\theta_1 - \sin^2 2\theta}$$

$$= \frac{\mu^2\cos^2\theta_1 + \cos^2\theta}{\mu^2\cos^2\theta_1 - \cos^2\theta},$$

or finally

$$\cos\delta_2 = \frac{(\mu^2+1)-(\mu^4+1)\sin^2\theta}{(\mu^2-1)-(\mu^4-1)\sin^2\theta} \qquad \text{.................(10)}.$$

As special cases we have

$$\text{for } \sin\theta = \frac{1}{\mu};\ \delta_2 = \pi,$$

$$\text{for } \theta = \frac{\pi}{2};\ \delta_2 = 0.$$

The difference in phase of the two components is best obtained directly by taking the real part of r_n/r_p which is equal to $\cos(\delta_1-\delta_2)$.

But $$\frac{r_n}{r_p} = \frac{\sin(\theta_1-\theta)}{\sin(\theta_1+\theta)} \cdot \frac{\tan(\theta_1+\theta)}{\tan(\theta_1-\theta)} = \frac{\cos(\theta_1-\theta)}{\cos(\theta_1+\theta)}$$

$$= \frac{\cos\theta\cos\theta_1 + \sin\theta\sin\theta_1}{\cos\theta\cos\theta_1 - \sin\theta\sin\theta_1}.$$

As the only imaginary quantity is $\cos\theta_1$, the expression is of the form

$$\frac{p+iq}{-p+iq},$$

the real part of which is $(q^2-p^2)/(q^2+p^2)$.

Hence

$$\cos(\delta_1 - \delta_2) = \frac{\cos^2\theta\cos^2\theta_1 + \sin^2\theta\sin^2\theta_1}{\cos^2\theta\cos^2\theta_1 - \sin^2\theta\sin^2\theta_1}$$

$$= \frac{1 + 2\mu^2\sin^4\theta - (1+\mu^2)\sin^2\theta}{1-(1+\mu^2)\sin^2\theta} \quad \ldots\ldots\ldots(11).$$

Equations (9) and (10), though giving us the values of the changes of phase, are unable to distinguish between an acceleration or retardation, and equation (11) does not tell us which of the two vibrations is ahead of the other. This ambiguity cannot be solved by the mere transformation of Fresnel's formulae. We may, it is true, show by means of (11) that $\delta_1 - \delta_2$ does not pass through zero, and hence reason that if δ_1 is positive, δ_2 must be negative, but recourse must be had to the complete dynamical theory in order to decide which component is accelerated. Though the subject has often been treated by various writers, it was only in 1884 that Lord Kelvin* pointed out for the first time that it is the vibration in the plane of incidence which is retarded, while the normal vibration is accelerated, and also that the difference of phase with the materials at our disposal, is always an obtuse angle. The latter conclusion may be derived from equation (11) as it is readily shown that the numerator within the range of total reflexion is positive and the denominator negative.

At incipient total reflexion (where $\mu\sin\theta = 1$) and for grazing incidence, there is a phase difference of 180°. Between these two limits of θ there is one angle for which the difference in phase is least. This angle is obtained from (11) by putting the differential coefficient of the right-hand side with respect to $\sin^2\theta$ equal to zero. This gives

$$(\mu^2+1)\sin^2\theta = 2\,;$$

the corresponding minimum retardation or maximum value of $\cos(\delta_1 - \delta_2)$ is

$$\cos(\delta_1 - \delta_2) = -\frac{6\mu^2 - \mu^4 - 1}{(\mu^2+1)^2} \quad \ldots\ldots\ldots\ldots\ldots\ldots(12).$$

An important practical application of these results was made by Fresnel. If it were possible to make the right-hand side of (11) equal to 0, there would be a phase difference of a right angle, which, if the original light was polarized at an angle of 45°, so as to make both components equal, would give circularly polarized light (Art. 6). Among the media at our disposal, there is none with a refractive index sufficiently high to give a difference of phase as small as $\pi/2$, but we can secure circularly polarized light by means of two successive reflexions, if $\delta_1 - \delta_2 = 3\pi/4$.

* *Baltimore Lectures*, p. 400.

Equation (11) may then be written

$$2\mu^2 \sin^4\theta - (1+\sqrt{\tfrac{1}{2}})\,\{(1+\mu^2)\sin^2\theta - 1\} = 0.$$

This is a quadratic equation which may be solved, and has in general two roots. Thus for glass of refractive index 1·5, 1·55 and 1·6, the following table gives the calculated values of the two solutions θ_1 and θ_2.

TABLE II.

$\mu =$	1·5	1·55	1·6
$\delta_1 - \delta_2 = 135°$	50° 14′ 53° 13′·5	45° 14′ 5 57° 5′	44° 21′ 56° 41′

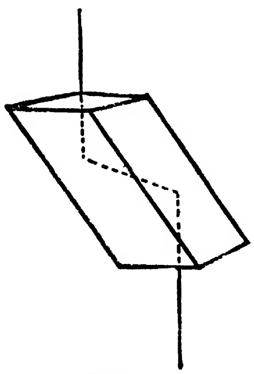


Fig. 34.

Fresnel's rhomb is a rhomb of glass (Fig. 34) which gives circularly polarized light after two total reflexions in the manner described.

Of the two possible angles for the rhomb, the larger is chosen, because it gives a smaller error for slight changes in refrangibility or deviations from the theoretically correct incidence.

CHAPTER IV.

THE PRINCIPLE OF SUPERPOSITION.

29. The Interference of Light. Huygens drew attention to the observation that the passage of a beam of light through an aperture is in no way affected by the passage of another beam through the same aperture. As he pointed out, different people may look at different objects through the same opening without noticing any blurring due to the overlapping of the large number of waves which must pass through the opening. The waves cross each other at the aperture without in the least interfering with each other's course.

We explain this independence of the separate waves by the principle of superposition (Art. 15) according to which the combined effect of a number of displacements may be obtained by adding the separate displacements, taking account of direction as well as magnitude. The principle of superposition may be applied also to velocities and accelerations but not to the squares of any of these quantities. In Art. 4 it has been shown that two periodic motions of the same frequency, of amplitudes a_1, a_2 and phase difference δ, combine to form a periodic motion having an amplitude the square of which is

$$a_1^2 + a_2^2 + 2a_1 a_2 \cos \delta.$$

The principle of superposition applied to the squares of amplitudes would account for the two first terms only and would therefore give erroneous results.

The illumination of a surface on which light falls depends on the square of the amplitude and in a most important group of optical phenomena, the results seem therefore to contradict the principle of superposition. It is indeed found that two rays of light may neutralize each other's effects so that darkness results where there was light when each ray produced its separate illumination. This effect has been called "interference of light." The term is a convenient one because it allows us to group together a certain class of phenomena but it is important to realise that interference is a direct result of the principle of superposition. If "interference" is to retain its position in Physics as a convenient term it becomes necessary to define it clearly and to adhere strictly to the definition. I shall adopt the following:

If the observed illumination of a surface due to two or more pencils of light is not equal to the sum of the illuminations due to the separate pencils, we say that the pencils have interfered with each other and class the phenomenon as one of "interference."

The distinctive feature of the definition is the inclusion of the adjective "observed." It means that the effects must be observed directly and not only be capable of being made visible by some additional appliance such as a spectroscope. The importance of this point will appear when we discuss special cases.

30. Calculation of the combined effects due to two separate sources. Let P and Q (Fig. 35) represent two particles which are sending out waves, the motion at P and Q being simply periodic. Let the vibrations be in a direction perpendicular to the plane of the diagram and identical as regards amplitude and period, and let the phases at P and Q be the same. Consider a point S on a distant screen, the plane of which is parallel to the line PQ and perpendicular to the plane of the diagram.

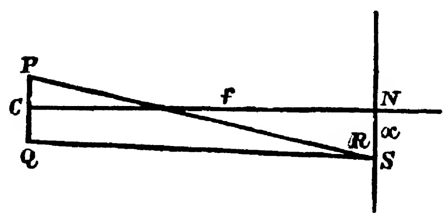


Fig. 35.

The two motions produced by P and Q at S, considered as acting separately, are parallel to one another, since they are both perpendicular to the plane of the diagram, and they have also approximately the same intensity, if the distance of the screen from the two sources is great compared with the distance of the two sources from one another. There will be a difference of phase between the two vibrations due to difference of the distances PS and QS. If R be a point on PS, such that $PR = QS$, the phase at R of the vibration transmitted along PS, must be the same as the phase at S due to the vibration transmitted along QS. Hence the difference of phase between the two vibrations at S, will be $\frac{2\pi}{\lambda}(PS - QS)$.

Let C be the middle point of PQ, and from C draw CN perpendicular to the plane of the screen, and cutting it in the point N.

Let $CN = f$, $NS = x$, and $PQ = c$.

Then
$$PS^2 = f^2 + \left(x + \frac{c}{2}\right)^2 \qquad \text{.........................(1)},$$
$$QS^2 = f^2 + \left(x - \frac{c}{2}\right)^2 \qquad \text{.........................(2)}.$$

Hence $$PS^2 - QS^2 = 2cx.$$

Therefore $$PS - QS = \frac{2cx}{PS + QS}.$$

If x is small compared to f, we may write $2f$ instead of $PS + QS$, the error committed being of the order of magnitude x^2/f^2.

The difference of phase between the two vibrations at S is therefore

$$\frac{2\pi}{\lambda} \cdot \frac{xc}{f}.$$

Let a denote the amplitude which would be produced at S by each source acting separately.

Then the resultant amplitude at S, due to both sources, is

$$2a \cos\left(\frac{\pi}{\lambda} \cdot \frac{xc}{f}\right) \quad \text{..............................(3).}$$

The amplitude is variable and depends on the angle x/f.

Thus considering points situated on the line NS on the screen, the point N for which $x = 0$, is a point of maximum intensity.

The intensity at points on either side of N diminishes symmetrically, and becomes zero when $x = \pm \frac{\lambda f}{2c}$. After this the intensity increases and reaches a maximum again when $x = \pm 2 \cdot \frac{\lambda f}{2c}$.

The points of maximum intensity are at equal distances $\frac{\lambda f}{c}$ apart and the points of minimum or zero intensity lie halfway between the points of maximum illumination.

So far only those points have been considered which lie in the plane PQN, but there is no difficulty in including points outside the central plane. If a point T be taken vertically over S, and at a distance z from it,

$$PT^2 = PS^2 + z^2$$

$$QT^2 = QS^2 + z^2$$

$$\text{....................}$$

$$PT^2 - QT^2 = PS^2 - QS^2.$$

As long as PT is, to the approximation required, equal to PS, *i.e.* as long as z^2 is neglected,

$$PT - QT = PS - QS.$$

Hence the illumination at T is the same as the illumination at S, and the illumination of a screen placed at NS consists therefore of a system of alternately bright and dark rectilinear bands, which are at right angles to the plane PQN.

If the distance of the screen is altered, the distance of the bands diminishes in direct proportion to the distance from the source, and all bands for which the difference in optical length $PS - QS$ is the same, lie in a plane at right angles to the plane of the paper, and passing through CS.

These results require some qualification as they depend on the squares of x and z being neglected. The complete investigation is not, however, difficult. The locus of the surfaces of equal difference in phase is determined by the condition that $PS - QS$ is a constant, a condition which defines the surfaces as hyperboloids of two sheets having P and Q as foci. The intersections of these hyperboloids with the plane of the screen are hyperbolas, and not straight lines, as found by the approximate method, but when the distance between P and Q is small, and those bands only are considered which are situated near the centre of the screen, the hyperbolas are very slightly curved, and may, near the central plane, be considered to be straight lines.

If the two sources of light are not in the same phase, but vibrate with a difference of phase which remains constant, the interference bands are formed as before, but the whole system is shifted to one side.

Let the vibration emitted from the source P be represented by $a \cos\left(2\pi \frac{t}{T} + \alpha\right)$ and that from Q by $a \cos 2\pi \frac{t}{T}$. Then at a point S on the screen, the difference of phase, which before was $\frac{2\pi (PS - QS)}{\lambda}$ will now be $\frac{2\pi (PS - QS)}{\lambda} + \alpha$ and the position of the bands will be given by the equation

$$PS - QS = n\lambda - \frac{\alpha\lambda}{2\pi}.$$

The bands are still at the same distance $\lambda f/c$ apart, but the whole system is displaced sideways by an amount equal to $\lambda f\alpha/2\pi c$. The assumption that the oscillations at P and Q are perpendicular to the plane PQN, may be removed, provided that these oscillations are parallel to each other, for under experimental conditions the distance PQ is so small compared with the distance of the screen, that the inclination between the displacements at S caused by parallel disturbances at P and Q may be neglected.

31. Conditions necessary for the experimental illustration of interference. Two homogeneous sources radiating from two points near each other, would, according to the last paragraph, produce a pattern of unequal illumination on a screen, though the position of the bands of maximum and minimum illumination could not *à priori* be determined unless we knew the differences in phase between the oscillations of the sources. It is observed that interference does not take place when the two rays are derived from two independent sources and this is the best proof we have that no available source of light emits homogeneous light. Quasi-homogeneous light requires for its analytical representation the overlapping of a number of waves having closely adjoining wave-lengths λ_1, λ_2, etc. If we pick out a particular wave-length λ_n and combine it with the same wave-length coming from the other source, we should obtain on the screen an interference pattern having a position on the screen which depends on the relative phases. Another wave-length would in general give the same pattern in a different position. The superposition of all wave-lengths would produce an average uniform illumination of the screen. Hence:

"*Independent sources of light even when emitting quasi-homogeneous light do not give rise to interference effects.*"

The experimental conditions of interference are obtained by deriving the oscillations originally from the same source. Two centres of radiation emitting vibrations which are related in phase owing to their having originated at the same ultimate source are said to be "coherent."

32. Young's experiment. Both on account of its historical importance and the simplicity of its arrangement, Young's experiment deserves the first place. Two small apertures P and Q (Fig. 36) were illuminated by light which originally had passed through another aperture at O. After passing through P and Q, the waves spread out in all directions, and falling on the screen SS', produce equally spaced interference bands. If P and Q are equidistant from O, the phase at P and Q will be the same, hence the central band will be at N. The equality of phase at P and Q holds for waves of all frequencies, and therefore the experimental conditions of Art. 31 are realized, and (3) correctly represents the distribution of amplitudes. As the distance between the bands depends on the wave-lengths, the light should be nearly homogeneous, if it is desired to observe the effects under the simplest conditions. A great number of bands may thus be seen. To give an idea of the scale on which the experiment has to be conducted, we may take as an example, the distance between P and Q to be 1 mm. and the distance of the screen from the aperture to be 1 metre. The distance between the bands is then for red light

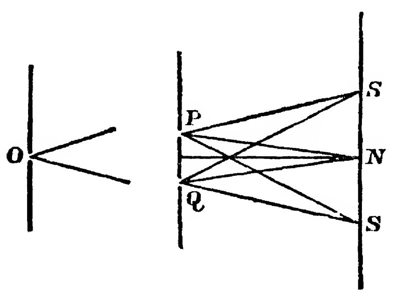


Fig. 36.

$$\frac{\lambda f}{c} = \frac{6 \times 10^{-5} \times 100}{\cdot 1} = \cdot 06 \text{ cms.}$$

and similarly for blue light ·04 cms. The bands are therefore very close together; if we wish to space them further apart, either the distance of the screen has to be increased, or the apertures have to be put closer together.

33. Fresnel's experiments. In Fresnel's celebrated experiments, the two dependent sources were secured by forming two vertical images of a narrow illuminated slit.

Fresnel's Mirrors. In the first of the two methods to be described, two inclined mirrors were used to obtain the vertical images.

In Fig. 37 OM_1 and OM_2 represent two plane mirrors, which have their planes at right angles to the plane of the diagram. Two images A and B are formed by reflexion of the light coming from S.

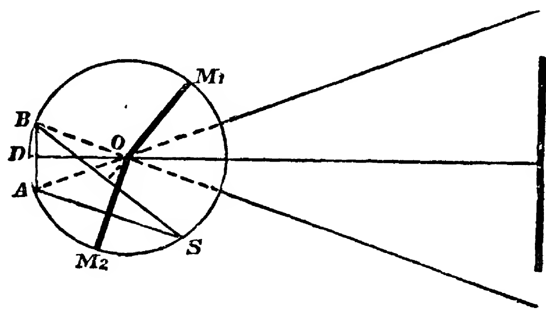


Fig. 37.

The distance between the two images depends upon the angle of inclination of the two mirrors. Let D be the middle point of AB and let DO be produced to meet a distant screen in C. Then C will be the centre of the system of interference bands, formed upon the screen. To calculate the distance between the two images A and B, we note that A being the image of S formed by the plane mirror OM_2, the distance of A to any point on OM_2 is the same as the distance from S to the same point. Hence $OS = OA$. Similarly $OS = OB$. Hence the points A, B and S lie upon a circle with centre at O.

Hence $$\angle AOB = 2\angle ASB$$
$$= 2\omega,$$

where ω is the angle between the two mirrors.

Therefore $$\angle BOD = \omega.$$

Now let $$OS = b \text{ and } OC = a.$$

Then $$DO = b\cos\omega \text{ and } DC = a + b\cos\omega.$$

Also $$AB = 2BD = 2b\sin\omega.$$

The distance between the bands produced on the screen by two sources of light, has been proved to be $\lambda f/c$.

In this case $$f = DC = a + b\cos\omega,$$
and $$c = AB = 2b\sin\omega.$$

Therefore the distance between the bands produced by Fresnel's mirrors is $\dfrac{\lambda(a + b\cos\omega)}{2b\sin\omega}$ or, since ω is a small angle, $\dfrac{\lambda(a+b)}{2b\omega}$.

Fresnel's Biprism. In the second method, the two images are obtained by doubling a single source by means of refraction. Suppose two similar small-angled prisms OPR, OQR are placed base to base as in the figure. This constitutes what is termed Fresnel's Biprism. If a source S is placed symmetrically behind the two prisms, two virtual images of it are formed say at A and B.

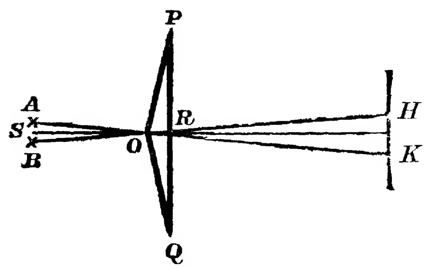


Fig. 38.

To calculate the distance between the bands, we make use of the fact that a prism of small angle α deviates any ray which falls on one of the faces in a direction nearly normal to it by a quantity $(\mu - 1)\alpha$, where α is the angle of the prism. Hence the vertical images of the slit are at the same distance from the prism as the object, and if b be the distance of the slit from the prism, $2(\mu - 1)b\alpha$ measures the distance between the vertical images. If a be the distance of the screen from the slit, the general expression for the distance between the bands reduces to

$$\frac{\lambda a}{2b(\mu - 1)\alpha}.$$

It should be noticed that the distance between the vertical images in this case, which represents the distance between the two sources of light producing interference, depends on the refractive index, and therefore on the wave-length. Plate I. Fig. 1 is a photograph of the interference bands formed by Fresnel's biprism. The rhythmic variation in the intensity of the bands is due to a diffraction effect which will be further alluded to in Art. 36.

34. Subjective method of observing interference bands. When interference phenomena are observed on a screen in the manner described, the bands are very close together, unless the screen is at a considerable distance from the sources, and in that case, a strong light has to be used if the bands are to be seen. There is, however, no difficulty in magnifying the bands by optical means. It has been shown in Art. 22 that the optical distance between object and image formed by a lens, is the same when measured along all rays. If therefore the screen be removed, and the rays crossing at any point P be focussed by a lens on another screen, the difference in phase between the two rays at the geometrical image of P is the same as the difference in phase at P. The interference pattern on the second screen will therefore be an image of the interference pattern on the original screen. If the lens in this argument is represented by the focussing arrangement of the eye, so that the retina represents the second screen, the interference effects will be seen just as if they were projected on a screen coincident with the plane for which the eye is adjusted. We can also interpose between the eye and the plane for which the eye is focussed a magnifying glass or eye-piece, and this enables us to measure the distance between the bands, for we may introduce a movable cross-wire in the focal plane of the eye-piece. This is practically Fresnel's arrangement, and the one which is generally adopted.

If the interference bands are observed through a telescope focussed for infinity, the interference pattern at the focus of the telescope is the image of that which would be formed at infinity, were the telescope away.

The simplest mode of seeing Young's interference bands has been described by Lord Rayleigh*. Two plates of glass are silvered; a fine line is ruled on one of them, and two fine parallel lines, as close together as possible, on the other. The ruling of the lines removes the silver film, so that we have now two opaque plates, one containing one slit, and the other, two slits close together. If the double slit is placed close up to the eye, and the other a short distance from it, interference bands are seen when the two plates are so adjusted that their slits are nearly parallel. The whole arrangement is easily constructed, and can be mounted in a tube.

35. Observations with quasi-homogeneous and white light. Our calculations based on the ideal case of homogeneous light are, up to a certain point, applicable to such quasi-homogeneous radiations as are at our command, because the position of the interference fringes belonging to two nearly equal wave-lengths λ_1 and λ_2 occupy nearly coincident positions, when the difference in path of the two interfering rays is only a few wave-lengths. As the difference in path is increased the fringes cease to be coincident and therefore appear less distinctly until ultimately the interference effects cease altogether. To illustrate this with the help of the formulae already obtained, consider two values of λ_1 and λ_2 to be the extreme wave-lengths emitted by a quasi-homogeneous source. The distance of the nth band (excluding the central one) from the centre is, with the previous notation, $n\lambda_1 f/c$, and so long as the difference between $n\lambda_1$ and $n\lambda_2$ is only a fraction of the

* *British Association Report*, 1893, and *Collected Works*, Vol. IV. p. 76.

wave-length, the fringes are practically overlapping, but when that difference amounts to half a wave-length the bright bands due to λ_1 are coincident with the lines of zero illumination belonging to λ_2 and the fringes cease to be visible. An interesting effect occurs when fringes are formed by sodium light which emits two sets of quasi-homogeneous radiations (the two sodium lines) having wave-lengths respectively in the neighbourhood of two values of λ_1 and λ_2 differing by the thousandth part of their value. In this case therefore when n is 500, $n(\lambda_1 - \lambda_2)$ is equal to half a wave-length and the two interference patterns dovetail into each other in such a manner that the fringes become very indistinct or cease to be observable. But when n is 1000, or any multiple thereof, $n(\lambda_1 - \lambda_2)$ is a multiple of a wave-length and the fringes again occupy identical positions. The fringes therefore show a periodic variation as the difference in path increases, becoming alternately distinct and indistinct and this continues until the want of homogeneity of each radiation separately destroys the interference effects.

The investigation of quasi-homogeneous radiation might easily be extended to white light but it is interesting to vary the point of view and to treat white light as a series of impulses in the manner explained in Art. 20. Let therefore two identical impulses start simultaneously from P and Q (Fig. 35). It is obvious that they will reach the screen simultaneously only at the point N. At every other point S they will follow each other at an interval which is the greater, the greater the distance NS. As there can be no interference except by overlapping we must conclude that no interference effects take place with white light. This seems to be contradicted by observation and an instructive insight into the phenomena of interference is gained by solving this apparent antagonism between two views which we know to be analytically identical. When observing interference effects with such white light as is at our command we notice a few coloured bands somewhat resembling the fringes observed with quasi-homogeneous light but less distinct. There is undoubtedly here an interference effect, but all we have a right to assert is that the interference is on the retina or rather in our physiological sensation. We can draw no conclusion on the presence or absence of interference on the screen on which we apparently see the interference pattern. To make this point clear imagine for a moment that the disturbance is a mechanical one which can set a pendulum placed at S into motion. The swing of the pendulum after the two impulses have passed over it depends on the relation of the time interval between them and the periodic time of the pendulum. If the interval is equal to the period the resulting motion of the pendulum will be a maximum, if the interval is half a period the effect of the first blow will be neutralized by the second. In the latter case therefore there will be *interference* of the two disturbances, but the interference is in the pendulum and not in the medium through which the disturbance has passed. The essential point to note is that if an instantaneous impulse produces an effect on the pendulum which is lasting, two successive impulses can mutually interfere, because they are now able to add their effects. The case of disturbances producing luminous sensations in our eyes is analogous. The fact that our retina is more sensitive to oscillations of certain periods and that different periods produce different sensations proves that there is some mechanism which responds to an impulse just as the pendulum responds to a blow. The effect of an impulse is in both cases a directed motion which does not disappear instantaneously and successive impulses are therefore capable of interfering with each other. The same argument holds if the interference fringes are received on a photographic plate. The interference here is due to the mechanism which makes the plate more sensitive to certain radiations than to others. To free ourselves from these actions we must determine the energy conveyed by the combined radiations and this can be done by means of a bolometer or thermopile. If the duration of each impulse were indefinitely short the thermopile would indicate no interference with Fresnel's mirrors or any other arrangement capable of showing interference with homogeneous light. The instantaneous impulse may be looked upon as an ideal white light: an extreme case which just as the ideal homogeneous wave, is not realised in nature. The light such as we find it emitted by a luminous body though called white, changes in character with the temperature and gives some preference to certain radiations over others. To represent this quasi-white light we must, as has already been pointed out, give the impulse such form and duration as will make the distribution of energy along the spectrum agree with observation. Its shape then approaches that illustrated by Fig. 10 and the distance which it covers may be taken to be approximately three wave-lengths of red light, becoming shorter with rise of temperature. If we are still allowed to call a disturbance of this kind an impulse it may be seen that in close proximity to N (Fig. 30) the two impulses overlap and here therefore some interference may take place, but there are no fringes, *i.e.* there is no rhythmic variation of intensity. If instead of observing with the eye or a photographic plate we were to measure the heating effect of the interference pattern of a double source, we should find a diminution of intensity from the centre outwards, reaching a minimum when the retardation is about half the wave-length at which the white light has its maximum of intensity. The illumination then increases to a maximum after which it very slightly falls off and remains constant*.

If we place the slit of a spectroscope at a point S of the screen, where with eye observations no interference is noticed in white light, the spectrum may be found to be traversed by interference fringes. In the light of our definition of interference this must be interpreted to mean that the interference is produced in the spectroscope. We shall indeed find when we come to discuss the theory of gratings that they convert each impulse into a series of impulses so that when two separate impulses follow each other their effects in the spectroscope may overlap with the result that they can interfere. At present we may deal with the effects observed when a spectroscope is used to examine an interference in white light by decomposing the light into its homogeneous constituents. Equation (3) gives for a given distance x from the centre the resultant amplitude, the illumination is consequently proportional to

$$\cos^2\left(\frac{\pi}{\lambda}\frac{xc}{f}\right).$$

According to the value of λ this varies between 0 and 1, and if the spectroscope separates the constituents the spectrum is seen to be crossed by bright and dark bands. If N represents the wave number, *i.e.* the number of waves per centimetre, a bright band appears when

$$N = \frac{nf}{xc},$$

n being an integer number. The bands are therefore equally spaced on a scale of wave numbers. If N_1 and N_2 are two wave numbers at which bright bands appear

$$N_2 - N_1 = \frac{mf}{xc}$$

and

$$m = \frac{N_2 - N_1}{N_1} n,$$

where N_2 would define the mth band counting from N_1.

By counting the number of bands in a given part of the spectrum between wave-lengths which are respectively equal to $1/N_1$ and $1/N_2$ we may therefore determine the number n, where n/N_1 is the difference in path between the two interfering rays.

36. Difficulty of illustrating simple interference phenomena by experiment. The simple mathematical treatment of the interference phenomena which we have so far studied, neglects certain

* *Phil Mag.* Vol. XXXVII. p. 541. (1894.)

effects which disturb the simplicity of the experimental verification. Thus the biprism of Fresnel (Fig. 38) shows interference only in the angle *HOK*, but a wave diverging from *A* and limited at *O*, the extreme geometrical boundary being *OK*, is not propagated entirely like a complete spherical wave. Certain so-called diffraction effects, which will have to be discussed in detail, take place; these alter the distribution of light, especially in the neighbourhood of the extreme rays *OH* and *OK*, and there appears a rhythmic variation in the brightness of the fringes, which sometimes makes their measurement difficult. The bands seen in Fresnel's mirrors are subject to the same irregularity, owing to the limitation of the beams by the rays *AO* and *BO*. Young's arrangement is free from this particular defect, but suffers from another. The slits at *P* and *Q* do not radiate light equally in all directions, but the intensity is a maximum in the directions *OP* and *OQ* respectively, and there are some directions (Art. 53) in which the light is totally absent. Hence here also, though from a different cause, the experiments give a rhythmic variation in the intensity of the interference fringes, which affects to some extent the positions of the maxima. We are therefore led to look in another direction for experimental methods to show interference in its simplest form.

37. Light incident on a plane parallel plate. When light is incident on a plane parallel plate, images of the source are formed by reflexion at the two surfaces; the reflected and transmitted beams may then show interference effects due to the overlapping of the waves coming from these images.

Let *LM* and *L'M'* (Fig. 39) be the parallel surfaces of a transparent plate, and *AB* an incident plane wave-front, which gives rise to a reflected wave *CD* and a refracted wave *RS*. This refracted wave will be reflected internally so as to be parallel to *R'S'* and however many internal reflexions take place all wave-fronts inside the plate are equally inclined to the surfaces and must be either parallel to *RS* or to *R'S'*. Similarly all waves which pass out of the surface *LM* must be parallel to *CD* and all those passing out of the surface *L'M'* must be parallel to *A'B'* which is parallel to *AB*. We have therefore in the reflected and transmitted beam a series of wave-fronts following each other at regular intervals. The plate may be looked upon as being a resonator which sends the incident energy to a greater or less extent in the forward or backward direction according to the relation between that interval and the wave-length. The result will be a flux of light inside the plate normal to *RS* and *R'S'* respectively. We shall neglect any absorption of light in the glass plate, so that the sum of the intensities of the reflected and transmitted beams must equal the intensity of the incident beam.

The first step in the calculation must consist in obtaining the differences in phase of the different coincident wave-fronts. Making

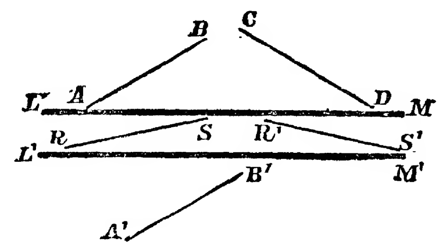


Fig. 39.

use of the fact that we may calculate optical distances between two wave-fronts along any ray connecting them, we may take some one wave-front AB in the incident beam (Fig. 40) and some one ray BS,

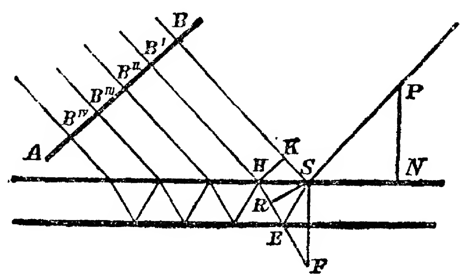


Fig. 40.

which at S is reflected towards P. Tracing the ray PS backwards through the plate, we find a ray $B'H$ such that starting from the original wave-front AB, it coincides with SP after one internal reflexion at E and refraction at S. The phase at P of the wave to which this ray belongs is determined by its phase at B, and the optical length of $B'HESP$. Similarly we may obtain a number of rays through B'', B''', B^{iv}, which will coincide along SP having been reflected two, three and four times, at the lower surface of the plate.

The difference in optical length between the two first rays is the same as the difference in optical length between $HE+ES$ and KS, K being the foot of the perpendicular from H on BS. If RS is drawn at right angles to HE, the optical length of KS is the same as that of HR. This follows from the fact that HK and RS are parallel respectively to the incident and reflected wave-fronts. The difference in optical length is now that due to the path $RE+ES$, or drawing the normal to the plate through S and producing HE to F, its point of intersection with the normal; the difference in optical length is μRF where μ is the refractive index of the plate. Noting that SF is twice the thickness of the plate (e), and that the angle at F is the angle between the refracted ray HF and the normal to the plate, for which we may write γ, it is finally found that the difference in optical length is $2\mu e \cos\gamma$. To obtain the difference in phase at P, we must, however, take account of the fact that the reflexions may be accompanied by change of phase, and we have already shown (Art. 25) that according to the principle of reversibility, there must be a change of two right angles at either internal or external reflexion.

A difference of phase of two right angles is equivalent to the addition to the optical length of a quantity equal to half the wave-length λ measured in vacuo. The difference in path is therefore finally

$$2\mu e \cos\gamma + \frac{\lambda}{2}.$$

If we only considered the superposition of the wave which is reflected externally on the one reflected once internally, we should find that the intensity of the reflected wave would be a minimum whenever

$$2\mu e \cos\gamma + \frac{\lambda}{2} = \frac{m\lambda}{2},$$

m being an odd number, or by transposing, when

$$2\mu e \cos\gamma = n\lambda,$$

n being any integer. It will be noticed that the difference in path becomes less, as γ, and therefore also the inclination of the incident beam, increases.

Before discussing the bearing of this equation, we extend the investigation so as to include multiple reflexions.

We take the vibration at S, due to the incident light, to be represented by $\cos\omega t$, for which, according to Art. 8, we write $e^{i\omega t}$, rejecting at the end of the investigation the imaginary part. The vibration at S in the reflected wave may then be written $re^{i\omega t}$, where r is real, if there is no change of phase. An incident wave of unit amplitude would then be reflected as a wave of amplitude r.

We may similarly apply coefficients t to the waves which are transmitted from the outside to the inside, t' for waves transmitted from inside to outside, and s for waves reflected internally. A change of phase would be indicated by the coefficients ceasing to be real.

Taking account of the fact that each of the rays in Fig. (40) has passed through a distance which is longer than the preceding one by the same quantity, of which we have already found the optical equivalent to be $2\mu e \cos\gamma$, the corresponding difference in phase is

$$\delta = (4\pi\mu e \cos\gamma)/\lambda.$$

Hence we may write for the vibration at S of that ray which has been once reflected internally $stt'e^{i(\omega t-\delta)}$, and for the ray reflected internally three times, the expression becomes $s^3tt'e^{i(\omega t-2\delta)}$, so that the factor of $e^{i\omega t}$ in the complete effect at S becomes

$$r + stt'\left(e^{-i\delta} + s^2e^{-2i\delta} + s^4e^{-3i\delta} + \ldots\ldots\right).$$

The terms of the geometric series in brackets converge towards zero, and may be added up. We thus find for the amplitude

$$r + stt'\frac{e^{-i\delta}}{1 - s^2e^{-i\delta}} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(4).$$

Dealing similarly with the transmitted waves, the successive vibrations at E, on emergence, are found to be $tt'e^{i(\omega t-\epsilon)}$; $s^2tt'e^{i(\omega t-\epsilon-\delta)}$ etc. if ϵ is the difference in phase between S and E, so that the factor of $e^{i(\omega t-\epsilon)}$ in the resultant vibration becomes

$$tt'\left(1 + s^2e^{-i\delta} + s^4e^{-2i\delta} + \ldots\ldots\right),$$

or

$$tt'\frac{1}{1 - s^2e^{-i\delta}}.$$

Experiments show that we are justified in assuming that reflexion and refraction at the surface of transparent bodies involve no change of phase, except, in certain cases, a change of π which is equivalent to a reversal of sign of the amplitude. We may then apply the relations found in Art. 25 for this case, *i.e.*

$$r + s = 0,$$
$$tt' + r^2 = 1.$$

The expressions for the reflected and transmitted beams then become for the reflected wave

$$r\frac{1 - e^{-i\delta}}{1 - r^2e^{-i\delta}}e^{i\omega t},$$

and for the transmitted wave,

$$\frac{1 - r^2}{1 - r^2e^{-i\delta}}e^{i(\omega t-\epsilon)}.$$

The intensities I_r and I_t for the reflected and transmitted waves are obtained by the rule given in Art. 8:

$$\left.\begin{aligned} I_r &= \frac{4r^2\sin^2\frac{\delta}{2}}{1 + r^4 - 2r^2\cos\delta} = \frac{4r^2\sin^2\frac{\delta}{2}}{(1 - r^2)^2 + 4r^2\sin^2\frac{\delta}{2}} \\ I_t &= \frac{(1 - r^2)^2}{1 + r^4 - 2r^2\cos\delta} = \frac{(1 - r^2)^2}{(1 - r^2)^2 + 4r^2\sin^2\frac{\delta}{2}} \end{aligned}\right\} \ldots\ldots\ldots(5),$$

from which it follows, as expected, that

$$I_r + I_t = 1.$$

The reflected and transmitted beams are therefore always complementary. In the reflected beams, the intensity is zero whenever

$$\sin\frac{\delta}{2} = 0,$$

i.e. when

$$2\mu e\cos\gamma = m\lambda \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(6).$$

If the incident light consist of a number of parallel pencils such as we receive, *e.g.*, from the sky, the reflected and transmitted light in different directions will depend on the angle of incidence, the angle of refraction entering into the expression for δ. When the reflecting power is small so that r^4 is negligible, the expression for the intensities reduces to the form:
$$I_r = 4r^2 \sin^2 \frac{\delta}{2}; \; I_t = 1 - 4r^2 \sin^2 \frac{\delta}{2}.$$

In the reflected light, the distribution of intensities in different directions then follows the same law as that with which we are familiar in the case of the interference of two equal sources (Art. 30), while in the transmitted beam, white light is superposed so that the relative changes of intensities are small. The result is altogether different when the reflecting power is large. Fixing our attention on the transmitted light, we may write for the reciprocal of the intensity:
$$1 + \frac{4r^2}{(1-r^2)^2} \sin^2 \frac{\delta}{2}.$$
For glass and nearly normal incidence, r^2 is approximately $\cdot 05$, and the factor of $\sin^2 \frac{\delta}{2}$ amounts only to $\cdot 011$. The difference between the maximum and minimum intensity is therefore only about one per cent. But if the reflecting power is $\cdot 9$, which is equal to that of silver, the factor becomes 360, so that, at the minimum, less than one-third per cent. of the incident light is transmitted. An examination of the above expression will moreover show that near the incidence for which $\sin \frac{\delta}{2} = 0$, the intensity diminishes very rapidly. This implies that in the transmitted beam almost all the light is concentrated in directions very near to those for which the light is a maximum.

38. Colours of thin films. Although there is no difference in principle between the interference effects observed with thin films or with thick plates, the method of observation most favourable in one case is not suitable in the other, and it is therefore convenient to treat each separately. We consider first the case of thin films.

Let an eye focussed on S be placed at P (Fig. 40), and let the incident light be derived from a distant extended source. The rays which enter the pupil have all passed through S but belong to waves which originally fell on the plate at different angles, and therefore have different intensities when they reach S. The relative retardation of successive rays being $2\mu e \cos \gamma$, the difference of retardation produced by a small change $d\gamma$ of γ is $2\mu e \sin \gamma \delta\gamma$, where γ is the angle of incidence. It may easily be calculated that, when e is not more than about thirty wave-lengths and we allow for $\delta\gamma$ the range which is necessary to fill the pupil with light when the eye is placed about a metre from the film, the change in retardation is only a small fraction of a wave-length and the unequal inclination of the beams does not cause any serious change in the appearance. We may therefore discuss the effects by considering only the central ray SP entering the eye. The interference fringes depending only on the angle of inclination are circular rings having as centre, N, the foot of the perpendicular drawn from the eye at P, to the surface of the plate.

The directions of maximum illumination depending on μ and therefore on the wave-length, colour effects appear in white light and we explain in this manner the colours of soap bubbles, of thin layers of oily matter and other thin films. In the mathematical discussion of these colour effects we may use the results of the investigation of Art. 35, substituting for the retardation (xc/f) the one applicable to the present case. If the reflected light were examined by the spectroscope the number $(m-1)$ of bands between two wave-lengths determined by their numbers N_1 and N_2 is therefore

$$m = (N_2 - N_1)\, 2\mu e \cos\gamma.$$

If we take for wave-lengths at the limits of the visible spectrum $6{\cdot}5 \times 10^{-5}$ and 4×10^{-5}, m becomes equal to $2\mu e \cos\gamma \times 10^4$. If the refractive index be that of water and $\gamma = 45°$, the number is approximately equal to $19000e$. If brilliant colours are to appear to the naked eye there should not be more than two dark bands in the spectrum. Putting $m = 2$ we find that e must not be more than 10^{-4}, which is about equal to twice the wave-length of green light. Hence brilliant colours are only seen when the thickness of the film is of the order of magnitude of a wave-length. As the retardation becomes less with increasing angle of incidence, the brilliancy of the colour is greater with oblique incidence.

39. Fringes observed with thick plates. When the plates are thick, nearly homogeneous light must be used. Taking PS (Fig. 40) to be one of the incident rays, it is necessary for complete interference that all the rays passing through B, B', B'', etc. should be brought into coincidence, or at any rate, all those which have sufficient amplitude to contribute appreciably to the effect. When the plate is thick, these rays may be too wide apart to enter the pupil together, and in that case, observations must be made through a telescope. If, on the other hand, SP is considered to be the reflected ray, it follows that the incident wave must have a sufficient width. Hence it is necessary for satisfactory observations to have either a wide incident beam, or to collect a sufficient number of rays in the reflected beam.

The complete wave-front AB gives rise to a number of rays parallel to SP, which have the same phase along a plane drawn through P and normal to SP. The eye should therefore be adjusted for infinity, or a telescope used. Fringes produced by the reflexion of light at the two surfaces of a plane parallel plate of thickness considerably greater than the wave-length of light, were first observed by Haidinger, but we owe their investigation more particularly to Mascart* and Lummer†. If a plate of glass AB (Fig. 41), a few millimetres thick, and carefully worked so as to have its faces plane and parallel, be illuminated by a sodium flame at S, and the rays are reflected from a transparent plate CD, so as to strike the plate AB in a nearly normal direction, the waves partly entering AB and partly reflected at the front surface, will cause overlapping wave-fronts to leave the plate. After traversing CD, these rays may be made to enter an eye adjusted for infinity. Rings are then observed having for centre the point which is the foot of the perpendicular drawn from the nodal point of the eye to the plate. If it is not desired to observe the complete rings, the plate CD may be dispensed with, and the flame placed near the eye. If the light reflected from the plate is then directly observed, it is found to be traversed by curved fringes. For greater brightness, we may use Lummer's arrangement, in which CD is replaced by a concave perforated mirror such as that used by oculists. The rings are observed through the aperture at the centre of the mirror.

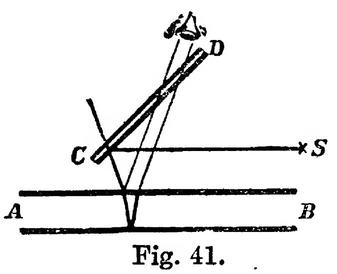


Fig. 41.

The condition for extinction is as before

$$2\mu e \cos\gamma = m\lambda,$$

where m is some integer and γ the angle of refraction in the plate.

If θ be the angle of incidence, and h the distance of the eye from the plate, the radius of the ring is

$$h \sin\theta = \mu h \sin\gamma.$$

This leads to the following construction (Fig. 42). On the radius OA of a circle take a point K_1 such that $\dfrac{OK_1}{OA} = \dfrac{m_0\lambda}{2\mu e}$, where m_0 is the highest integer which gives to the fraction on the right-hand side a value less than one. From K_1 mark off towards O, equidistant points, K_2, ... K_n, the equal distance being $\lambda OA/2\mu e$. From K_1, K_2, etc. draw perpendiculars to OA meeting the circle in S_1, S_2.... Then on a

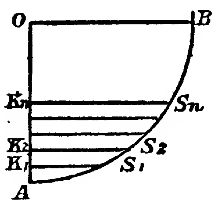


Fig. 42.

* *Ann. Chim. Phys.* XXIII. p. 16.

† *Wied. Ann.* XXIII. p. 49.

scale in which OA represents μh, the diameters of successive dark rings from the centre outwards are given by K_1S_1; K_2S_2; etc.

The number of rings is finite, and equal to the highest integer number which is less than OA/K_1K_2. The centre of the system of rings may be bright or dark according to the thickness of the plate and the refrangibility of the light used.

Fig. 2 Plate I. is a photograph of Haidinger's fringes obtained with a glass plate 3·6 mm. thick. The source of light was a small mercury lamp, the ultra-violet rays being absorbed by a solution of quinine. The actinic light was in consequence very homogeneous, being almost exclusively due to a violet mercury line.

40. Michelson's combination of mirrors. A very powerful optical combination for obtaining interference fringes was devised by Prof. Albert Michelson*. In principle, it is identical with the system which has been discussed in the previous article.

Let *s*, Fig. 43, be a source of light sending out waves towards a

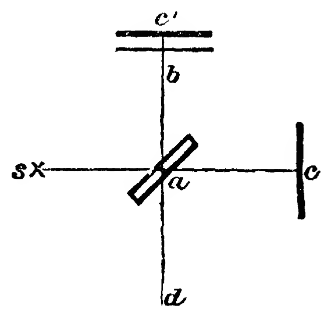


Fig. 43.

plate of glass *a* inclined at an angle of 45° to the wave-front. The mirror is lightly silvered at the front, so that the incident light is divided into two approximately equal portions, one being transmitted towards a mirror *c* and the other reflected towards another mirror *b*. At both these points reflexion takes place which sends the rays back towards *a*. Here once more there is partial reflexion and transmission, and two sets of wave-fronts will proceed from *a* towards *d*, one having passed over the course *sabad*, and the other over the course *sacad*. Neglecting for a moment the thickness of *a*, the optical length from *a* to *c* and back is the same as that from *a* to an imaginary mirror *c'* and back, *c'* being the image of *c* in the silvered surface of *a*. The interference effects observed are therefore identical with those of a plate having two reflecting surfaces *b* and *c'*. Compared with the arrangements previously described, Michelson's interferometer has the advantage that the passage of light between the two reflecting surfaces takes

* Described in a joint paper by A. A. Michelson and Ed. W. Morley, *Phil. Mag.* Vol. xxix. p. 449. (1887.)

place in air so that the difference in the refractive indices for waves of different wave-lengths is very small. The essential part of the arrangement is that two images of a source of light S are formed behind the mirror b; if e be the distance between b and c', the distance between the two images is $2e$.

The difference in optical length of two parallel rays formed by reflexion at b and c' respectively and brought to a focus is $2e \cos\gamma$ if γ denote the angle between their direction and the normal to the mirror. But as one of the rays pursuing the path *Sabad* has been reflected externally at the surface a, while the other travelling along ac and back has been reflected internally, there is an additional retardation of a wave-length of one with respect to the other (Art. 25). The total difference in optical length is independent of the position of S, which may be a source of light of finite dimensions. To an eye placed at d and focussed for infinity, the mirror will be seen to be covered with bright interference bands which are circular rings if all the adjustments are correct. Taking account of the change of phase at reflexion, the rings of minimum brightness are given by

$$2e \cos\gamma = m\lambda.$$

It is clear that the higher values of m must correspond with smaller values of γ, and that m must be smaller than $2e/\lambda$ in order that γ may have a real value. If m_0 be the highest integer compatible with that condition the angular radii of successive dark rings are equal to $\sin\gamma$.

The diameters of the rings are proportional to $\sin\gamma$, and

$$\sin^2\gamma = 2(1-\cos\gamma) = 2\left(1-\frac{m\lambda}{2e}\right) = 2\left(1-\frac{m_0\lambda}{2e}\right) + \frac{s\lambda}{e},$$

where $m = m_0 - s$. The first term is constant. By giving to s the successive values 1, 2, 3, etc. we obtain the values of $\sin^2\theta$ for successive minima. If the centre itself is a point of minimum light $m_0\lambda/2e = 1$ and the radii of successive dark rings are as the square roots of successive integer numbers, or what comes to the same thing, as the square roots of successive *even* integers. If the centre is a point of maximum light, we find similarly that successive dark rings are as the square roots of successive *odd* integers.

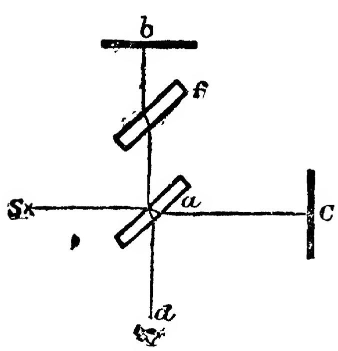


Fig. 44.

In practice, the thickness of the plate a has to be compensated. This is done by interposing an equal plate f (Fig. 44) into the path of the rays between a and b.

The complete theory of Michelson's combination of mirrors* should

* Bureau International des Poids et Mesures: *Travaux et Mémoires*, Vol. VII.

include the case of a slight inclination of the mirrors. It is sufficient here to point out that if b and c' (Fig. 43) are inclined to each other, the optical conditions are the same as those which will be discussed in Art. 41. When the distance of c is such that c' intersects b, there will be a dark band coinciding with the line of intersection and the fringes in the neighbourhood will be sensibly parallel to this line.

41. Newton's rings. The system of rings observed near the point of contact of a lens with a glass plate is one of the oldest of known interference phenomena. Its elementary theory is simple, though its complete investigation is troublesome, owing to the curvature of one of the boundaries at which light is reflected. The colours observed in Newton's rings are the colour of thin films, the film being the layer of air included between the lens and the plate on which the lens is placed. The characteristic distinction between Newton's rings and the phenomena we have already discussed is that the film has now a variable thickness.

The simplest case of a film of variable thickness would be presented by a transparent wedge (Fig. 45).

A plane wave AB falling on the wedge, we may select one ray $B'S$ reflected at S towards P, and another $B''H$, such that after refraction at H and reflexion at E, it meets the upper surface at the same point S. Owing to the inclination of the two surfaces, the refracted ray SP' is not now coincident with SP, though the inclination is small, if the angle of the wedge is small. The difference in optical length between the two rays is $2\mu e \cos\gamma$, or taking account of the change of phase at reflexion, $2\mu e \cos\gamma + \frac{\lambda}{2}$. In this expression,

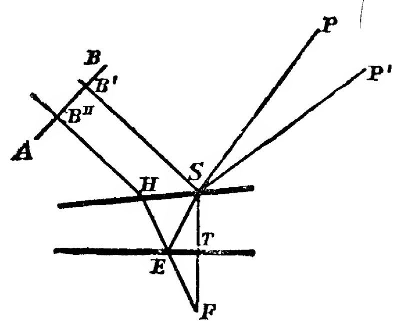


Fig. 45.

e denotes the length of the perpendicular from S to the lower surface of the plate (which may be taken to be the thickness of the plate at S) and γ is the angle of incidence on the *lower* surface. The inspection of the figure explains how the expression is derived. Neglecting all rays which have suffered more than one internal reflexion, an eye placed so as to receive both rays SP and SP' and focussed on S, will observe a maximum or minimum of light, according as $2\mu e \cos\gamma$ is an odd or even multiple of half the wave-length. If the source of light be extended, waves coming from different directions must be considered. Each of these waves supplies two interfering rays at S, and the difference in path depends to some extent on the inclination. Hence the eye focussed at S combines on the retina a number of rays which are not under identical conditions, and the interference will not be so simple or so complete as when the plate has equal thickness throughout. The disturbing effect will be small when the plate is thin, and may be neglected for the first few rings in Newton's experiment when the thickness does not exceed more than a few wave-lengths. For thicker plates, observations may be improved by reducing the aperture of the pupil by interposing a slit so as to narrow the pencil of light which can enter the eye. If the eye is focussed for a point different from S, interference is still observed, though with a slightly changed difference in path. This is shown by imagining the ray $B''H$ to be shifted either to the right or to the left. If it is shifted to the right, SP' moves to the right, and its intersection with SP moves away from P, so that the eye has now to focus for a more distant point. The inclination of SP' remains the same, but the length of the path inside the plate is longer or shorter according as BH'' moves to the right or left. It follows from what has been said that we may apply the equations of plane parallel plates to films of varying thickness, so long as their thickness is small. The interference is made more complete by restricting the source so that it only subtends a small angle at the film. If the incidence is nearly normal, a slight variation in the direction of the incident beam has very little effect on the difference in optical length of the two interfering waves, which also prevents confusion of the interference effects. Thus while in the case of the plane parallel plates previously considered, the colours are due to the varying inclination of the incident beam, the thickness of the plate being everywhere the same, we now confine ourselves to the same direction of incidence, and obtain the colours as a consequence of the changing thickness. Fig. 46 illustrates how Newton's rings may be observed under nearly normal incidence. The lens LL' being placed on the plate AB, an inclined transparent plate MN serves to reflect light coming from the source at S, while the eye observes the light reflected from the film included between the lens and the plate, and transmitted through MN. To calculate the diameter of the rings, it is only necessary to obtain a relation between the thickness at any point e, and the corresponding distance ρ from the point of contact; if R is the radius of curvature of the lower surface LOL' of the lens, the geometry of the circle gives

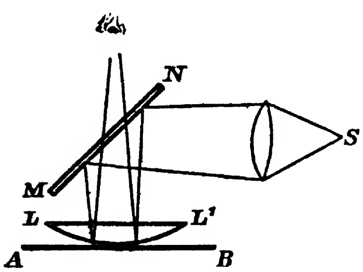


Fig. 46.

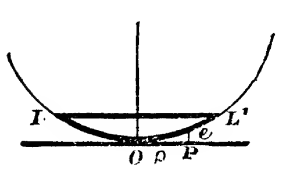


Fig. 47.

$$\rho^2 = e\,(2R - e)$$

or so long as e is small compared with R, so that its square may be neglected,

$$\rho^2 = 2eR.$$

The difference in optical length of the two interfering rays is $2e + \frac{\lambda}{2}$, if the observation is conducted so that the medium included between the lens and plate is air, so that $\mu = 1$. The diameters of the rings of maximum illumination are obtained by making $2e$ an odd multiple of half a wave-length, so that they are proportional to the square roots of successive *odd* numbers, while the dark rings will have diameters proportional to the square roots of successive *even* numbers. The centre of the ring system is dark, though black only when the lens and plate are of the same material, in which case the whole light is transmitted if there is optical contact at O.

The minimum of light at the centre of the system of rings appears as a consequence of the retardation $\lambda/2$ at internal or external reflexion, there being in consequence total destruction with no difference of path. If the upper and lower surfaces are made of different material, and the film has a refractive index intermediate between the two, the centre of the ring system on the other hand is bright, as the half wave retardation now disappears. Thomas Young showed this by introducing oil of sassafras between a lens of crown glass and a plate of flint glass.

In the transmitted system of Newton's rings, the colours are less brilliant. Their position is easily deduced from the fact that the effect at every point must be complementary.

Plate I. Figs. 3 and 4 represent photographs of Newton's rings. The same mercury lamp was used in both cases as the source of light, the rays producing the photographic effect being principally derived from one violet and one ultra-violet radiation. In Fig. 4, the ultra-violet radiation has been blocked out by an absorbing screen and hence the appearance is that due to practically homogeneous light. In Fig. 3 we may observe the effect of the overlapping of two systems of rings which alternately strengthen and neutralize each other. Where the dark and bright rings of the two systems coincide the rings are clearly defined; where the bright ring of one overlaps the dark ring of the other the rings are very indistinct. Similar effects may be observed with sodium light owing to the difference in the wave-lengths of the two components of the sodium doublet, but the two wave-lengths being more nearly equal the intervals between the regions of maximum definition are much greater.

42. Brewster's bands. When light passes through a plate of glass, a small change in optical length may be made by slightly inclining the plate. If it is desired to observe so-called interference due

PLATE I.

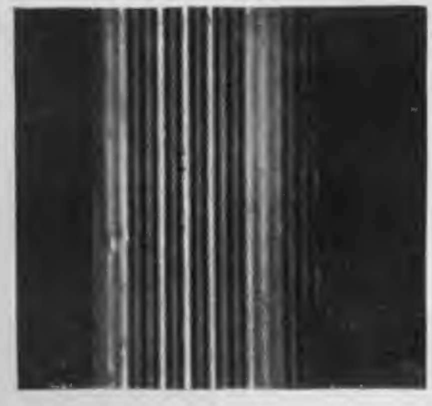


FIG. 1.

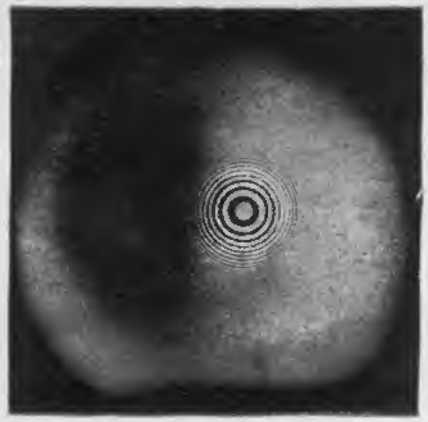


FIG. 2.

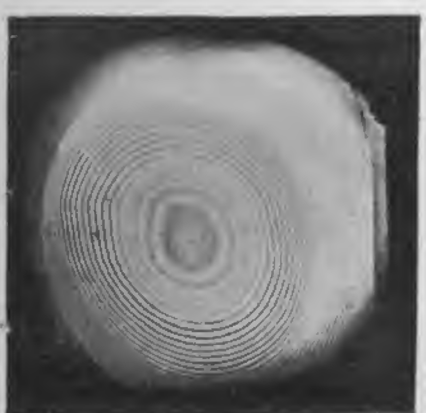


FIG. 3.

FIG. 4.

FIG. 5.

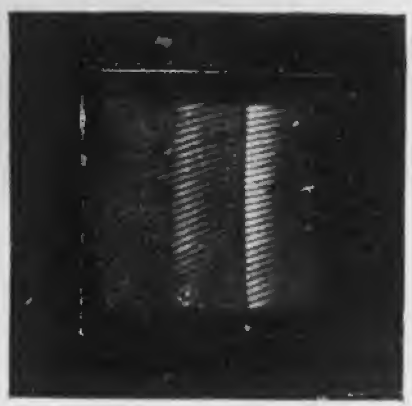


FIG. 6.

[*To face page* 78.

to this alteration, it is necessary to interpose an exactly equal plate into another portion of the same beam, so that in the first instance there may be equality of path, which is then slightly disturbed by the inclination of one of the plates. This leads to the following arrangement due to Brewster. A plate of glass, which should be as nearly as possible plane parallel, is cut in half so as to obtain two plates of equal thickness; one is slightly inclined to the other and light passed through them. The course of the rays which are brought to interference is shown in Fig. 48. If the plates were parallel, the optical lengths would be equal. A slight inclination of one causes a relative change of phase in the overlapping beams, and when an illuminated surface is viewed through the plates, coloured bands are seen to traverse the field. The interference fringes may also be observed in reflected light, and Fig. 49 shows how we may obtain a number of different sets of interfering rays according to the number of internal reflexions. In the first system, marked 1 in the figure, two rays are brought to interference, the first having been once reflected internally in the plate A, and the second once in plate B. The second system consists of three rays, one of which has been reflected once in each plate, and the two others twice in one plate and not at all in the other. The course of the third system is also shown in the figure, and the further ones need hardly be considered, as the intensity of light rapidly diminishes by multiple reflexions. To prevent confusion, it is necessary to place a screen at SS' to limit the incident beam. If the bands were observed near the plane of the figure they would be seen to be strongly curved, and the field of view would only contain bands formed by rays having large retardations. To find the position of the central band which is that in which the relative retardation is zero, we start from the fact that the optical length in each plate depends only on the angle of incidence of the light. The thickness and refractive index of the two plates being the same, the optical length is the same for all rays which in their passage from one plate to another are equally inclined to both plates. These rays all lie in a plane which is parallel to the line of intersection of the plates and normal to the plane bisecting the angle between A and B. The image of that plane in the plate B is the locus of the central band. In order to observe the fringes near the central

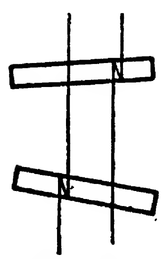


Fig. 48.

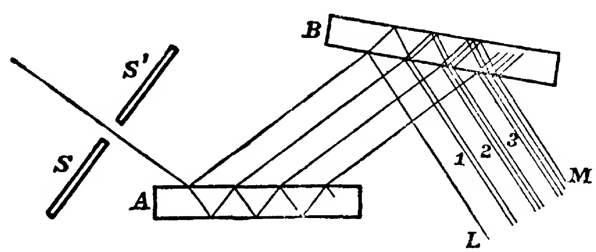


Fig. 49.

band it is necessary that the plane containing it should intersect the plate B, but that at the same time the line of view should not be obstructed by A. It is easy to see that the conditions are always difficult and often impossible to fulfil if the plates are vertical and the observations are carried out as usual in a horizontal plane. A convenient arrangement is to tilt the plates slightly round a horizontal axis. If this inclination is equal and opposite, the central band is horizontal, and can easily be observed from any position, such as M in the figure. When one of the plates is then moved round a vertical axis the bands remain horizontal, but are shifted upwards or downwards.

43. The Interferometer of Fabry and Perot. Fabry and Perot have constructed an instrument, capable of rendering important service, by utilizing the fringes observable in light transmitted through a plate which in this case consists of a layer of air formed between two parallel glass plates, the inner surfaces of which are thinly silvered.

We obtain the intensity of the transmitted beam by applying equation (5) of Art. 37; the value of μ which now represents the refractive index of air may be taken as equal to unity. The maximum intensity of the transmitted light takes place when

$$2e \cos \gamma = m\lambda.$$

This is exactly the same relation as holds in Michelson's interferometer, with the exception that the maxima and minima are interchanged, the inversion of phase at the reflexions being avoided. The appearance of the rings is however entirely different, in consequence of the multiple reflexions at the surfaces of the silvered air plates. The difference is an important one, and it tells in favour of the multiple reflexions. While with single reflexions the intensity gradually varies according to the sine law from the maxima to the minima and back to the maxima, multiple reflexions concentrate the light when the reflecting power is great, almost entirely near the maxima, the intensity falling off rapidly on either side. As the late Lord Rayleigh has pointed out, the space between the two reflecting surfaces acts like a resonator. When the bright rings are narrow, the separation of two radiations having nearly the same wave-length is assisted, but this narrowing cannot be pushed beyond a certain point on account of the absorption of light in the layers of silver through which it has to pass. Hence the deposit of silver has to be thin, and this diminishes its reflecting power. Nevertheless, considerable advantage is gained.

The interferometer is extremely valuable in the determination of the ratio of the wave-lengths of the spectroscopic lines, and the method of procedure will be explained in connexion with the general problem of the measurement of wave-lengths (Art. 70). For some purposes Fabry and Perot found it necessary to design their instrument so that the distance between the two plates was adjustable. This requires very accurate workmanship if the parallelism is to be maintained during the adjustment. The construction is easier when the plates are at fixed distances, and this suffices for most purposes.

43 A. Lummer plates. We have seen how the multiple reflexions in thick plates can be made use of to separate two homogeneous radiations of nearly identical period. In Fabry and Perot's arrangement the internal reflexions were strengthened by silvering the surfaces of the plates. Lummer achieved the same object by utilizing the high reflecting

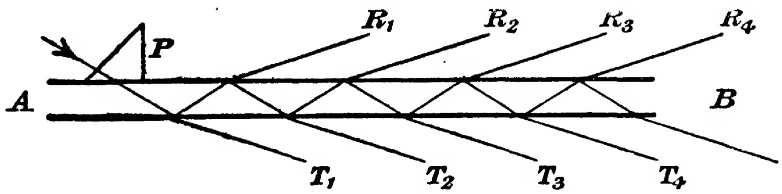


Fig. 50.

power near the angle of total internal reflexion. The light enters the plate AB through a small rectangular prism P cemented to the plate. The object of this prism is to reduce the loss of light at the entry into the system. The transmitted rays, T_1, T_2, T_3, if collected by a lens, combine together in exactly the same way as in Fabry and Perot's interferometer, and the same formulæ hold for their intensity. There is however an important difference in the reflected beam, in so far as the first reflexion is eliminated from the system R_1, R_2. In the expression (4), Art. 37, we have therefore to suppress r and (5) becomes

$$I_r = r^2 \frac{(1 - r^2)^2}{1 + r^4 - 2r^2 \cos \delta} = r^2 I_t.$$

The intensity of the reflected wave is therefore proportional, and, if the coefficient of reflexion be great, nearly equal to the transmitted wave.

44. Stationary vibrations. When two waves of the same amplitude and period proceed in opposite directions we may represent the displacement by

$$a \cos (\omega t - nx) + a \cos (\omega t + nx) = 2a \cos nx \cos \omega t.$$

The right-hand side of the equation shows that the phase is now constant everywhere, but the amplitude depends on x and is zero whenever x is an odd multiple of a quarter of a wave-length. The amplitude has a maximum value in the intermediate places at which x is a multiple of half a wave-length. The combined disturbance of two waves crossing each other in this way is called a stationary vibration. An alteration in the phase of one of the waves shifts the positions of the maxima and minima, but does not alter their distance. Altering *e.g.* the phase of the wave proceeding in the negative direction, by two right angles, we should get

$$a \cos (\omega t - nx) - a \cos (\omega t + nx) = 2a \sin \omega t \sin nx.$$

These stationary waves are easily illustrated in the case of sound waves. Experimental investigation in the case of light involves great difficulties, which were, however, successfully overcome by O. Wiener*, who succeeded in demonstrating the stationary vibrations formed at the surface of a mirror by stretching a sensitized film, the thickness of

* *Wied. Ann.* XL. p. 203 (1890).

which was only a fraction of a wave-length, obliquely across the wave-system formed close to the mirror.

If $A\,A_1$ be a silvered glass plate on which light falls normally and $B\,B_1$ a very thin sensitive collodion film (which of course must be attached to another glass plate not shown in the figure), the stationary vibrations will have their maxima at points a, b, c, d, such that their distances from the plate—neglecting possible changes of phase at reflexion—are multiples of half a wave-length. If a spectrum be projected on the plate such that the image of the slit is parallel to $A\,A_1$ each spectrum line is crossed by dark bands corresponding to the minima of intensity in the stationary vibration. This is shown in Fig. 6, Plate I, which is a reproduction of one of Wiener's photographs. The spectrum—with its violet end to the right—is that of the electric arc, and shows mainly two carbon bands, the H and K calcium lines being faintly seen between these bands. The inclination of the interference fringes to the spectrum lines in the photograph is due to a slight inclination of the slit. The success of Wiener's experiments depended on the formation of photographic films less one-twentieth of a wave-length thick, and he was able to show that the method is well adapted to investigate the behaviour of light at reflexion, such as the reversal of phase when it is reflected at the surface of a denser medium.

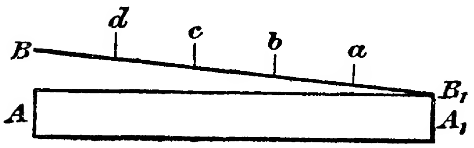


Fig. 50 *a*.

Drude and Nernst having succeeded in obtaining sufficiently thin fluorescent films, observed the stationary vibrations by their fluorescent effect.

Lippmann's Colour Photography is based on the formation of thin layers of reduced silver deposited within a photographic film, the layers being half a wave-length apart. They are formed by the stationary vibration of waves of light reflected from a surface of mercury over which the sensitive film has been extended. When viewed in reflected light the colours of thin plates are seen, and that colour shows a maximum of intensity which has a wave-length equal to twice the distance between the layers. We therefore see chiefly the colour belonging to the wave which originally had formed the stationary vibration. The great difficulties which stood in the way of success can be realized by observing that the nuclei of silver which form the layers must have linear dimensions not exceeding a small fraction of a wave-length. It took some years of patient work before a method was found that proved satisfactory. The first photographs which gave results imitating roughly the natural colours, were defective owing to the varying sensibility of the film when exposed to different parts of the spectrum, but perfect ortho-chromatism was obtained in 1893. The possibility of reproducing natural colours in the same fashion had already occurred to Wm. Zenker*, and to Lord Rayleigh†. The experimental realization due to Lippmann is, however, a very considerable experimental achievement.

* *Lehrbuch der Photochromie.* † *Collected Works*, Vol. III. p. 13.

45. Applications. We may divide the principal remaining applications of interference phenomena into two classes. In the first, a measurement of the difference in optical length of two paths is aimed at. Instruments used for this purpose have been called interference refractometers. Fresnel, already, in conjunction with Arago, made use of interference bands to measure the difference in the refractive indices of dry and moist air. Two parallel tubes, filled with the gases to be examined, were placed in the path of a plane wave-front which traversed the tubes longitudinally: the displacement in the bands observed when dry air was replaced by moist air served as a measure of the difference in refractive index. Jamin carried out important measurements in an apparatus in which use was made of Brewster's interference bands. The tubes containing the gases were placed in the paths of the two first of the rays reflected from the plate A in Fig. 49, and the differences in optical length showed themselves by a displacement of the bands. Simpler and more effective methods are now available, and we need not therefore enter into the details of Jamin's apparatus. But a useful little appliance used in the measurements may be described. This is a "compensator," consisting of two plates of glass (Fig. 51) capable of being rotated round a horizontal axis AB, and placed at such a distance from each other that each plate receives the light which has passed through one of the tubes. Rotation round the horizontal axis alters the thickness of glass traversed. The alteration being different for the two plates a measurable retardation of one set of rays, as compared with the other, is produced. If the central band, having been displaced by the change of pressure in the tube, is brought back by the compensator to its original position, the difference in refractive index between the air under partial exhaustion and the air at atmospheric pressure, can be measured. Different gases may be compared in a similar manner.

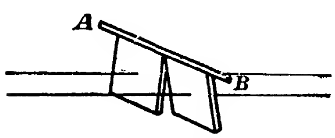


Fig. 51.

Lord Rayleigh's* form of Refractometer more nearly approaches the original instrument of Fresnel and Arago.

Light coming from a fine slit L and rendered parallel by a collimator lens C of 3 cms. aperture passes through two brass tubes side by side, and soldered together. These tubes, 20 cms. long and 6 mm. in bore, are closed at the ends by plates of worked glass, so connected as to obstruct as little as possible the passage of light immediately over the tubes. The light having passed through the tubes enters two slits and is brought to a focus F by means of a lens. The optical arrangement is practically identical with that which gives Young's fringes (Art. 32). The fringes are observed by means of an eyepiece. To secure better illumination and sufficient magnifying power, the eyepiece is cylindrical, so as only to magnify in a horizontal direction. It is made of a short length of glass rod, 4 mm. in diameter. There are two systems of bands, one formed by

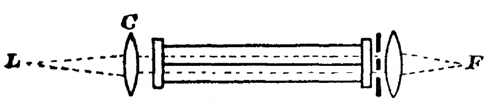


Fig. 52.

* *Collected Works*, Vol. IV. p. 364.

light which has traversed the gases within the tubes, the other by light which passes above them. If different gases are to be compared with each other, as regards their refracting power, their pressure is adjusted until the system of bands formed by light which has passed through the tubes is coincident with the system formed by the light which has passed above the tubes; the retardation in the two tubes is then the same. If the experiment be repeated at a different pressure, then the ratio of the changes of pressure for each gas is the inverse ratio of the refractivities $(\mu - 1)$ of the gases.

Other refractometers have been constructed, chiefly with a view to separating the path of the interfering rays laterally as much as possible, so as to leave more room for the tubes or other apparatus to be introduced into the path of the rays. It is sufficient to refer to the apparatus of Zehnder*. It should be noticed, however, that the lateral separation of the rays is by no means always an advantage. One of the experimental difficulties in delicate optical measurements consists in keeping the temperature sufficiently constant, or at any rate, not to introduce a *difference* in temperature into the two optical paths. The nearer these are together, the easier will equality of temperature be secured. Where a separation of rays is necessary or advisable for other reasons, Michelson's arrangement, which has already been described, will probably be found to be the most advantageous. The applications which Michelson has made with this instrument to the investigation of the constitution of nearly homogeneous radiation will be referred to in Arts. 68 and 193.

An appliance, useful in many optical measurements, is the "bi-plate" which serves either to separate or to bring together two parallel beams of light. It consists of two plane parallel plates of glass cemented together at an angle. Their action is sufficiently illustrated by Fig. 53.

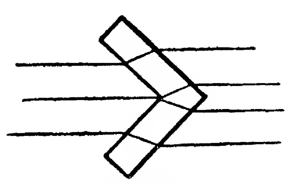


Fig. 53.

In the applications of the phenomena of interference which have been dealt with so far, the problems are of a purely optical nature. We turn now to the second class of applications in which optical methods are used for linear measurement.

Fizeau has used Newton's rings to examine the coefficients of expansion of certain substances. The body to be examined, cut *e.g.* into the form of a cube, is placed on a plate which, by means of screws passing through it, supports a lens. The upper surface of the cube is polished. If the lens be adjusted so as to leave a small air space between it and the cube, Newton's rings may be observed. If the whole arrangement is raised in temperature a change takes place in the rings which depends on the altered distance between the upper surface of the cube and the lens. Knowing the effect of temperature on the refractive index of air and the coefficient of dilatation of the other part of the apparatus, that of the cube may be deduced. Fizeau has measured in this manner the expansion of crystals in different directions. For a more detailed

* *Ztsch. f. Instrumentenkunde*, 1891, p. 275.

account of the apparatus and method of obtaining the result from the observed displacements of Newton's rings, Mascart's *Optics*, vol. 1, p. 503, may be consulted.

Perfectly flat surfaces are sometimes required in optical investigations, and it is a matter of great difficulty to work them so as to satisfy optical tests. Not the least of the difficulties consists in testing the surface when it is nearly flat, so as to discover where its faults are and how they may be corrected. The late Lord Rayleigh* used for this purpose the bands seen between a horizontal surface of water and the carefully levelled surface which is to be examined. The latter surface is supported horizontally at a distance of about one or two millimetres below that of the water. By the aid of screws the glass surface is brought into approximate parallelism with the water. When the surface is perfectly flat, the interference bands are straight, while a curvature of the bands always implies a curvature of the surface. In the paper referred to it is shown how to interpret the curvature of the surface by means of that of the bands. The chief difficulty in applying the method consists in securing perfect steadiness, so as to avoid the effects of tremor on the water surface.

45 A. Earth Tides. An important application of interference phenomena was made by Michelson † in his investigation of the tidal distortion of the earth's surface. The optical arrangement is shown diagrammatically in Fig. 53*a*. AA_1 is a steel tube 6 ins. in diameter and 100 ft. long, buried 6 ft. below the surface of the ground so as to eliminate the effects due to wind or change of temperature. The tube is hermetically sealed at one end and filled with water which stands at the level LL_1. The mirror b of the interferometer is placed just below the surface of the water, the two other mirrors being in the position indicated by c and a. If the earth's surface be distorted by tidal attraction, the distance between the water level and the mirror b slightly alters and the interference fringes shift. It was found that the tidal distortion of the earth's body as a whole produces a maximum shift between 10 or 20 fringes, which can be measured to one-tenth of a fringe, giving an average error of less than one per cent. By a suitable arrangement a continuous record can be obtained. Small local tremors produce no effect but seismic disturbances can be observed and measured. "It appears thus that the installation may serve as a seismograph; and it may even be possible to observe a slow secular change in the apparent level due to an inclination of rock strata the ultimate rupture of which causes the earthquake ‡."

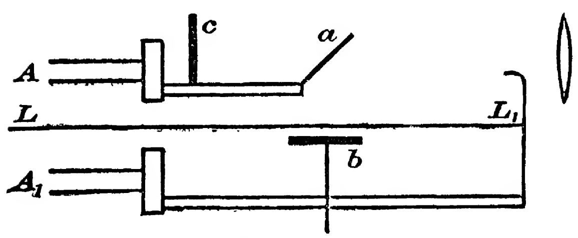


Fig. 53 *a*.

* *Collected Works*, vol. IV. p. 202.
† Guthrie Lecture, *Proc. Phys. Soc. of London*, Vol. XXXIII. p. 276.
‡ *loc. cit.* p. 280.

46. Historical. Christian Huygens (born April 14, 1629, at Haag in Holland, died June 8, 1695) is the founder of the undulatory theory of light. His treatise on light appeared in 1690, and contains the explanation of the reflexion and refraction of light by means of the principle which now bears his name. He also demonstrated how double refraction could be explained by means of wave-surfaces having two sheets, and in particular showed how, in Iceland Spar, a wave-surface consisting of a sphere and spheroid accounted for the laws of refraction of both rays.

Sir Isaac Newton (born Jan. 5, 1643, at Grantham in Lincolnshire, died March 21, 1727) did not favour the wave theory of light. He was misled by the apparent difference in the behaviour of waves of sound which, after passing through an opening, spread out in all directions, and the rays of light which pass in nearly straight lines. This seemed a formidable difficulty, and Huygens' attempts at explaining the apparently rectilinear propagation of light were not clear or convincing. While there is no doubt that Newton's great authority kept back the progress of the undulatory theory for more than a century, this is more than compensated by the fact that the science of Optics owes the scientific foundation of its experimental investigation in great part to him. His experiments on the prismatic decomposition of white light do not fall within the range of this volume, but the phenomena of Newton's rings have been referred to. Newton discovered that the radii of bright or dark rings were determined by the thickness of the layer of air interposed, and found the correct law connecting the diameters of successive rings.

Thomas Young, born June 13, 1773, at Milverton (Somerset), studied medicine in London, Edinburgh and Göttingen. He was Professor of Physics at the Royal Institution in London between 1801 and 1804, but gave up his position in order to devote himself to the practice of medicine. He died on May 10, 1829. To Young belongs the merit of having been the first to state clearly the principle of the superposition of waves and to show how interference may be explained by means of it. Owing to the historical importance of this principle, on which the development of the undulatory theory of light entirely depends, the passage in which Young first introduced it may be quoted. It occurs in a paper read before the Royal Society on November 12, 1801, and runs as follows:

"PROPOSITION VIII. When two Undulations, from different Origins, coincide either perfectly or very nearly in Direction, their joint effect is a Combination of the Motions belonging to each."

"Since every particle of the medium is affected by each undulation, wherever the directions coincide, the undulations can proceed no otherwise, than by uniting their motions, so that the joint motion may be the sum or difference of the separate motions, accordingly as similar or dissimilar parts of the undulations are coincident."

Young's arrangement for observing interference fringes, which has been discussed in Art. 32, is thus described in his published lectures (1807):

"In order that the effects of two portions of light may be thus combined, it is necessary that they be derived from the same origin, and that they arrive at the same point by different paths, in directions not much deviating from each other. This deviation may be produced in one or both of the portions by diffraction, by reflection, by refraction, or by any of these effects combined; but the simplest case appears to be, when a beam of homogeneous light falls on a screen in which there are two very small holes or slits, which may be considered as centres of divergence, from whence the light is diffracted in every direction. In this case, when the two newly formed beams are received on a surface placed so as to intercept them, their light is divided by dark stripes into portions nearly equal, but becoming wider as the surface is more remote from the apertures, so as to subtend very nearly equal angles from the apertures at all distances, and wider also in the same proportion as the apertures are closer to each other. The middle of the two portions is always light, and the bright stripes on each side are at such distances that the light, coming to them from one of the apertures, must have passed through a longer space than that which comes from the other, by an interval which is equal to the breadth of one, two, three, or more of the supposed undulations, while the intervening dark spaces correspond to a difference of half a supposed undulation, of one and a half, of two and a half, or more."

There is no other reference to these experiments in Young's published paper, so that we do not know the size of the apertures, or their distance apart. Young seems to have attached more importance to the cases where the openings are wide and the intervening space narrow, though the theory of these cases is more complicated. Young was very successful in his explanation of the colour of thin films, especially in the mechanical analogy which he brought to bear on the change of phase which takes place at one of the reflexions. The otherwise formidable difficulty of explaining the dark centre of Newton's rings was thus at once satisfactorily overcome.

Fresnel's important work belongs more particularly to the next chapter. As regards simple interference and its experimental illustration, we owe him the method of inclined mirrors and of the biprism. He also showed how fringes could be observed subjectively through an eyepiece; a method of observation which enabled him to carry out accurate measurements.

Gabriel Lippmann (1845—1921) combined exceptional independence of thought with great perseverance and remarkable experimental powers. He was born in Luxembourg, but his parents soon settled in Paris, where he spent the remainder of his life. His first important contribution to Science was the invention of the capillary electrometer, which has attained a well-established position in physical and physiological laboratories.

His process of colour photography is described in the text, and the invention of the coelostat, which keeps the image of a finite part of the sky stationary for a lengthened period, is highly valued by astronomers.

CHAPTER V.

THE DIFFRACTION OF LIGHT.

47. Huygens' principle. By means of Huygens' principle, we may obtain the effect of a wave-front WF at a point P (Fig. 54), by dividing its surface into a number of elements, and adding up their effects. Our problem then consists in finding the law according to which a small portion of the surface may be supposed to act. If we consider the element at S to be an independently vibrating source, it is seen that its effect at P can only depend on the length of the vector SP, the angle which that vector forms with the normal to the surface, and the angle between the same vector and the direction of vibration at S. If the investigation be limited to homogeneous vibrations, we may obtain in a simple manner an expression for the displacement at P which yields, at any rate, one possible solution of the problem.

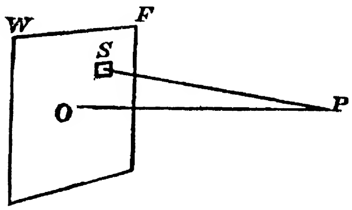


Fig. 54.

Draw PO, the normal to the wave-front WF, and call O the pole of P. Draw two circles with the pole as centre and radii OS and OR. The area of the ring included between these circles is

$$\pi(OS^2 - OR^2) = \pi(PS^2 - PR^2)$$
$$= \pi(PS - PR)(PS + PR).$$

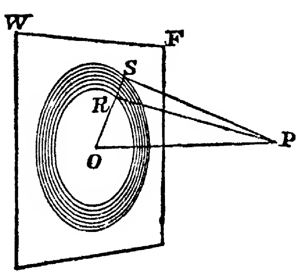


Fig. 55.

If $PS - PR = \delta$, and δ is a small quantity, the square of which can be neglected compared to PO, the expression for the area of the ring becomes

$$2\pi\delta PR.$$

If the ring be further subdivided into a very large number of concentric circles having radii OR_1, OR_2, such that

$$PR_1 - PR = PR_2 - PR_1 = PR_3 - PR_2 = \text{etc.},$$

the successive rings have equal areas, and their separate effects at P must be equal in magnitude. To calculate their joint effect, we must take account of the difference in phase of the vibrations reaching P from each separate ring. If a diagram be drawn in which the effect of each ring is represented by a vector, these vectors will be of equal length and will succeed each other at equal angular intervals.

Hence according to Art. 5 the resultant amplitude is $a\dfrac{n \sin \alpha}{\alpha}$, where α is half the difference in phase between the first and last vibration. The product na represents what would be the amplitude at P, if every point of the ring were at the same distance from that point. Writing c for this product, we find that the complete ring causes a vibration at P, having an amplitude $\dfrac{\sin \alpha}{\alpha}c$. Its phase is the arithmetic mean between the phases due to the first and last ring. If $PS - PR = \lambda$, $\alpha = \pi$ and the amplitude at P is zero. If $PS - PR = \lambda/2$, the amplitude at P is $2c/\pi$.

Divide now the whole wave-front into zones, Fig. 56, by rings of radii OK_1, OK_2...... such that

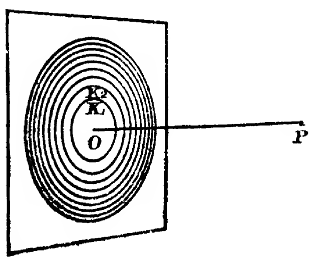


Fig. 56.

$$\frac{\lambda}{2} = PK_1 - PO = PK_2 - PK_1 = PK_3 - PK_2 = \text{etc.}$$

The resultant phase of two successive zones differs according to the above by two right angles, so that to obtain the total effects, we need only add up the effects of successive zones, giving the opposite sign to successive values.

Hence $$S = m_1 - m_2 + m_3 \ldots\ldots$$

where $m_1 = 2c_1/\pi$; $m_2 = 2c_2/\pi$ etc. The quantities c_1, c_2 etc. depend on the distance between each ring and P, and may also depend on the angle KPO or the angle between the direction of vibration and KP. These quantities all alter very little between one ring and the next and we may therefore take the difference between two successive values of m to be very small. This being so, a very simple expression for the sum of the series may be obtained.

Collecting the terms differently, the series may be written in the form

$$S = \frac{m_1}{2} + \left(\frac{m_1}{2} - m_2 + \frac{m_3}{2}\right) + \left(\frac{m_3}{2} - m_4 + \frac{m_5}{2}\right) + \ldots\ldots\ldots\ldots (1),$$

the last term being $\frac{1}{2}m_n$ or $\frac{1}{2}m_{n-1} - m_n$, according as n is odd or even. Each of the bracketed terms is small if the values of m alter slowly, but we should not be justified for this reason alone in neglecting them, because if their number is large, their sum may be comparable in magnitude to m_1. But assuming that the law of vibration is such that the effect of each zone is smaller than the arithmetic mean of the effects of the preceding and following zones, all terms in brackets are positive, and therefore

$$S > \frac{m_1}{2} \pm \frac{m_n}{2} \quad \text{..............................(2)},$$

where the plus or minus sign is chosen according as n is odd or even.

The series S can also be written in the form

$$S = m_1 - \frac{m_2}{2} - \left(\frac{m_2}{2} - m_3 + \frac{m_4}{2}\right) - \left(\frac{m_4}{2} - m_5 + \frac{m_6}{2}\right) + \text{......(3)},$$

and under the same conditions as before, each bracket is positive If m_1 is sensibly equal to m_2, and m_n sensibly equal to m_{n-1}, it follows that

$$S < \frac{m_1}{2} \pm \frac{m_n}{2}.$$

Comparing this with (2) it is seen that the bracketed terms are negligible, and hence

$$S = \frac{m_1}{2} \pm \frac{m_n}{2}.$$

The same conclusion is arrived at by supposing that the brackets in (1) and (3) are all negative. If a change in the sign of the brackets occurs in the course of the series, we may divide the series into two parts, and sum each part separately. We thus arrive again at the same conclusion that the whole effect is equal to one-half the sum of the effects of the first and last zones, unless the brackets in the expression (2) change so frequently in sign that the outstanding small effect at each reversal sum up to be an appreciable quantity.

Excluding such special cases, which need not be considered in any optical application, we may now apply our result to the calculation of the resultant effect of a plane wave-front extended but ultimately limited by a boundary which is not a circle having the pole as centre. In Fig. 57 the boundary is assumed to be square. We may draw all the circles complete until one of them touches the boundary. After that point is reached, parts of the zones are blocked out by the opaque screen, and the effect of these outer zones must gradually diminish and ultimately vanish. In this case, therefore, the effect of the last zone is zero, and we find that the resultant effect at P is equal in magnitude to half that of the first zone. Writing p for OP, the area of the central disc has been shown to be $\pi p\lambda$. To obtain its effect we must apply the factor $2/\pi$ and we thus find that it causes an amplitude at P which is the same as that produced by a surface of

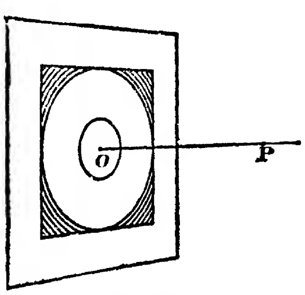


Fig. 57.

area $2p\lambda$, placed half-way between the centre and edge of the disc. If ks is the effect at P of a small surface s placed at O, the effect of the first zone is $2kp\lambda$, and the effect of the whole wave, as has been shown, is equal to that of half the first zone. The wave being plane, the amplitude at P is the same as at O. Calling that amplitude a, $kp\lambda = a$, and hence

$$k = \frac{a}{p\lambda}.$$

We must conclude that if a wave-front is split up into a number of small elements, we arrive at a correct result in the case of an extended plane wave of amplitude a, if we take the effect at a point P of a small surface s as regards amplitude to be $as/p\lambda$. The surface s is here supposed to be so small that the distances of its various points from P do not differ by more than a small fraction of the wave-length. The occurrence of p in the denominator can readily be understood, as the effect of an independent source on a point at a distance may be expected to be such that the intensity varies inversely as the square of the distance. If this be granted, it also follows that λ must occur in the denominator, as the factor of a must be of the dimensions of a number, and of the three quantities s, p, λ involving the unit of length s occurs in the numerator and p in the denominator.

It may now be shown that the value of k just obtained also gives correct results, when the wave-front is spherical. In Fig. 58 let waves diverge from a point Q and let it be required to calculate the effect at P from one of the wave-fronts WF. The only difference there can be between this problem and the previous one lies in the magnitude of the first zone, which must therefore be recalculated.

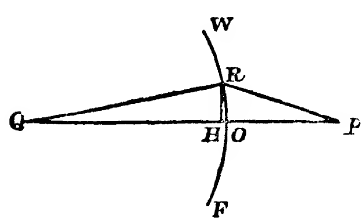


Fig. 58.

Let RH be drawn at right angles to PQ and let $QO = q$; $PO = p$; $RH = f$; $HO = t$.

Then neglecting powers of f higher than f^2

$$f^2 = 2qt,$$

also
$$f^2 = PR^2 - PH^2$$
$$= (PR + PH)(PR - PH),$$

and
$$PR = p + \frac{\lambda}{2},$$

$$\therefore f^2 = \left(\frac{\lambda}{2} - t\right) 2p.$$

Eliminating t,
$$f^2 = \frac{\lambda pq}{p+q}.$$

The effect of the first zone as regards amplitude is equal to $2ks/\pi$, where s is the surface of the zone. Substituting $s = \pi f^2$ and $k = a/p\lambda$, where a is the amplitude at O, the amplitude at P which is half the effect of the first zone is found to be $aq/(p+q)$ and varies therefore, as it should do, inversely as the distance from Q.

Returning to the case of plane waves we obtain another important result by considering the phase of the resultant vibration. The phase at P due to the action of any zone has been shown above to be half way between the phases due to portions of the zone which are respectively nearest to and furthest from P. Applying this to the central zone, the phase of the resultant vibration at P, if calculated in the usual way, should differ from that at O by

$$2\pi\left(p + \frac{\lambda}{4}\right)\bigg/\lambda \text{ or } \frac{2\pi p}{\lambda} + \frac{\pi}{2}.$$

But we know that the optical distance from O to P is simply p, and hence the difference in phase is $2\pi p/\lambda$. It follows that if we want to obtain the phase correctly at P by means of Huygens' principle, we must everywhere subtract a quarter of a wave-length from the optical distance, or imagine the wave-front to be shifted forward through that distance.

It may be well to recapitulate what it is that has been proved. An extended wave-front has been divided into zones, and grouped together in such a way that the effect of the whole wave could be shown to be equal to that of half the central zone which lies close to the pole O. The effect of a small surface s at a distant point P was expressed by ks, and it appeared that k was independent of the angle between the radius vector and the normal to the wave-front. But this only proves that at the central disc, the cosine of that angle being sensibly equal to unity, any effect of the inclination is eliminated. Similarly the result is independent of any possible effect of the direction of vibration.

The division of the wave-front into zones, drawn so that the distance of their successive edges from the point at which the amplitude of light is to be estimated, increases by half a wave-length, has rendered it possible to apply Huygens' principle in a simple and effective way. This mode of treating the propagation of waves being due to Fresnel, the zones should be called "Fresnel Zones."

48. Laminar zones. Instead of dividing the wave-front into circular zones, it is often more convenient to perform the division in a different manner. Let P (Fig. 59) be the point at which the light is to be estimated and WF the wave-front. Divide WF into a number of parallel strips at right angles to a central line HK. Let LM be such a strip, which may again be subdivided into smaller areas, chosen to be of such magnitudes that the resultant phases of two successive elementary areas are in opposite directions. If the strip be indefinitely extended in both directions, we may form a series as in the previous article, and find in this way that the total effect must be some definite fraction of that element of LM which is nearest to the central line HK. The whole effect being proportional to the width of the strip t, we may put it equal to kht, where k is the factor previously determined, and h some linear quantity. This expression asserts nothing more than that the effect of the strip is equal to that of an area situated in the central line HK, having a width t and a height h. The same reasoning may be applied to each of the strips which are parallel to LM, and we finally reduce the effect of the wave-front to that of a horizontal strip of width h. This may once more be subdivided. As the strip of width t produces an effect at P equal to kht, the effect of a strip of width h must be kh^2. Hence the effect of the complete wave-front is reduced to that of an area h^2 placed at O, O being the pole of P. If the amplitude is a, $kh^2 = a$. Hence

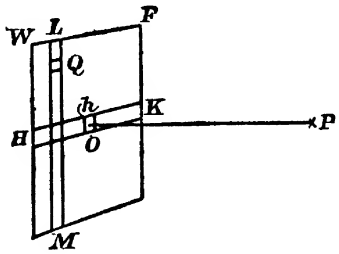


Fig. 59.

$$h = \sqrt{\frac{a}{k}} = \sqrt{p\lambda},$$

p being the distance OP; the effect as regards amplitude of a strip such as LM of width t is therefore $ta/\sqrt{p\lambda}$.

To obtain the resultant phase due to each strip, we make use of the previously established fact that in applying Huygens' principle, we obtain the optical distance by taking away a quarter of a wave-length from the actual distance between the source and the point at which the amplitude is required. We imagine therefore the whole wave-front to be brought nearer through that distance. Now the process of attaining the final resultant from the rectangular strips consists of two exactly equal steps, the first in obtaining the intermediate resultant of each vertical strip such as LM, and the second in summing up for the horizontal strip HK which represents that intermediate resultant. If the total effect of the two steps as regards phase, is to bring back the wave-front to its proper position, each step must contribute equally, and therefore the optical distance of each strip is obtained by taking away $\lambda/8$ from the actual distance. When the wave-front is divided into strips, it follows therefore that for the calculation of phases, we must imagine each strip to be brought nearer by $\lambda/8$. Or for simplicity of calculation we may say that we may take the optical distance of a strip to be equal to its actual distance, if we correct the final result by subtracting $\lambda/8$ from the calculated optical distance or 45° from the calculated phase.

We may now determine the widths, t, of the strips, so that their resultant effects at some given point are alternately in opposite directions. Let Q be the point and OT_1, T_1T_2, T_2T_3, etc. (Fig. 60) represent the widths. The total resultant effect of all the vertical strips has been shown to correspond to an optical distance of $p+\frac{\lambda}{8}$ if $OQ=p$, and if the resultant phases of successive strips are in opposite directions, the resultant optical distance of the nth strip must be:

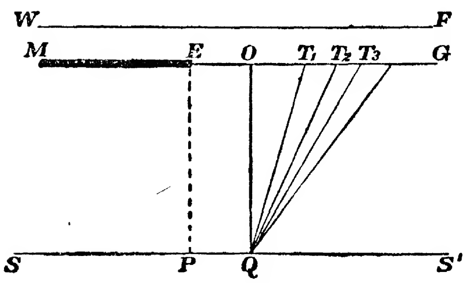


Fig. 60.

$$p+\frac{4n-3}{8}\lambda.$$

The phase at Q of the vibration due to any one strip $T_n\,T_{n+1}$ which is not near to O may be taken to be that belonging to the optical distance which is equal to the arithmetical mean of QT_n and QT_{n+1}. This may be seen by subdividing each strip into minor equal strips and assuming that the distances of successive subdivisions increase uniformly. This assumption is justified with greater and greater rigour the greater the angle between QT_n and QO; the error which is introduced in the strips which lie near O is found to be small and even for the second strip, $T_1\,T_2$, it may here be neglected. To obtain the right value for the resultant phase of each strip after the first, we must now put

$$QT_1=p+\tfrac{3}{8}\lambda,$$
$$QT_2=p+\tfrac{7}{8}\lambda,$$
$$QT_3=p+\tfrac{11}{8}\lambda \text{ etc.}$$

Note that the distance QT_1 has been derived from the consideration of the second strip and that we have not assumed that the phase of the effect of the first strip OM corresponds to the arithmetical mean of the distance of its edges. This would not have been correct because the distance between Q and the line OG passes through a minimum at O, and if the first strip were subdivided, we could not assume as we did for the other strips that the distances of the subdivisions increase uniformly. As regards phase we know however that for OT_1 it must be that corresponding to the optical distance $p+\frac{1}{8}\lambda$, because, by our construction, the phase of the disturbance due to each strip must be alternately in agreement with and in opposition to that of the resultant vibration, that of the first strip being in agreement.

To show the accuracy reached by the above simple reasoning I give the results of a more complete calculation of the distance QT_1 and QT_2 for which the error is greatest,

$$QT_1 = p + \tfrac{3}{8}\lambda - {\cdot}0046\,\lambda,$$
$$QT_2 = p + \tfrac{7}{8}\lambda + {\cdot}0016\,\lambda.$$

This number shows that for practical purposes the error introduced by the simplification we have made is negligible.

The width of successive strips is obtained from

$$OT_n = \sqrt{QT_n^2 - p^2} = \sqrt{\frac{4n-1}{4}\,p\lambda},$$

where λ^2 is neglected compared to $p\lambda$. Hence for the first strip

$$t_1 = \tfrac{1}{2}\sqrt{p\lambda}\,\sqrt{3},$$

for the second strip $\quad t_2 = \tfrac{1}{2}\sqrt{p\lambda}\,\{\sqrt{7} - \sqrt{3}\}$,

and generally

$$t_n = OT_n - OT_{n-1}$$
$$= \tfrac{1}{2}\sqrt{p\lambda}\,\{\sqrt{4n-1} - \sqrt{4n-5}\}.$$

The effect of the nth strip t_n is, as regards amplitude:

$$\frac{2}{\pi}\frac{t_n}{\sqrt{p\lambda}} = \frac{1}{\pi}\{\sqrt{4n-1} - \sqrt{4n-5}\}.$$

The numerical values of the effects are given in Table III. for $n=2$ to $n=12$. They have been calculated from the above expression, except for the first strip, for which the method fails to give correct results. The effect of this strip may be obtained by calculating the numerical value to which the series approaches, leaving out the first strip.

Remembering that the effect of the second strip is negative the series to be summed up is:

$$-\pi^{-1}[(\sqrt{7} - \sqrt{3}) - (\sqrt{11} - \sqrt{7}) + (\sqrt{15} - \sqrt{11})\ldots\ldots].$$

Its value is found to be $-{\cdot}1725$. As we know that the total effect of all the strips on one side of O must be $\cdot 5$, it follows that the effect of the first strip as regards amplitude must be $\cdot 6725$.

TABLE III.

Effects in Amplitude of Fresnel Strips.

No. of strip	Effect in amplitude	No. of strip	Effect in amplitude
1	+ ·6725	2	− ·2908
3	·2135	4	·1771
5	·1547	6	·1391
7	·1274	8	·1183
9	·1109	10	·1047
11	·0995	12	·0949

49. Preliminary discussion of problems in diffraction. When an obstacle is placed in the path of a wave-front and the shadow of the obstacle received on a screen, the boundary of the shadow is not sharp, but the light encroaches to some extent on the dark portions, while there are bright and dark fringes on the side towards the light. If we draw straight lines which proceeding from the source of light touch the shadow-throwing body, the intersections of these lines with the screen enclose what may be called the geometrical shadow, meaning thereby the shadow constructed according to the laws of geometrical optics. Owing to the fact that light consists of waves, the laws of geometrical optics are not strictly true, but the waves spread round the obstacle and encroach to some extent on the geometrical shadow. That they do not do so to a greater extent, was the principal difficulty of the wave theory in its earlier form. This bending round of the waves has been called the "Diffraction" of light. The simplest problems of Diffraction are those in which we imagine a plane or spherical wave to impinge on a plane perforated screen. Whatever form or position the apertures HK, $H'K'$ (Fig. 61) have, we can find the disturbance at a point P by Huygens' principle, if we know the disturbance at all points of the openings. In the usual solutions of the problems, the assumption is made that the disturbance is the same at all points in the plane of the screen as it would be if the screen were away. In other words, the screen simply obstructs the light which falls on its opaque portions, but does not otherwise alter the motion of the medium. That the assumption is one which needs justification may be understood by contemplating *e.g.* the flow of water through a pipe, in which the stream lines are parallel straight lines, and imagining that at some place a diaphragm is introduced across the pipe, leaving only an aperture much narrower than its cross section. We should here obviously arrive at erroneous results if we were to assume that the velocity of the water at all points of the opening has not been altered by the introduction of the diaphragm. In the case of the ordinary diffraction effects, it is found that the results arrived at by the simplified calculation are in agreement with experiment. This is a consequence of the small size of the length of a wave of light as compared with the other linear magnitudes which enter into the calculation, the errors introduced being sensible only within a few wave-lengths of the obstacle.

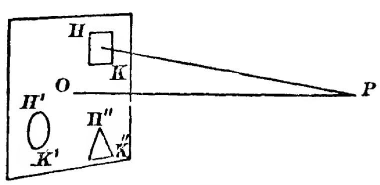


Fig. 61.

We are allowed therefore to use Huygens' principle in its simple form, provided we correctly introduce the contribution which each small surface element s at a point S of the opening contributes to the amplitude at P. If r be the distance PS, ϕ the angle between r an the perpendicular to the wave-front at S, and θ the angle between r and the direction of vibration, the effect for homogeneous vibration of a small surface s at P is according to Stokes:

$$\frac{s(1+\cos\phi)\sin\theta}{2r\lambda}.$$

This expression is based on the assumption that the *displacements* in the openings are everywhere the same as if the screen were away. Lord Rayleigh, on the other hand, has shown that if the *forces* acting across the plane of the screen are the same as if the screen were absent, the effect of s would be

$$\frac{s\sin\theta}{\lambda r},$$

and has also pointed out that so far as the treatment of diffraction problems is concerned, the terms depending on θ and ϕ disappear in consequence of interference, so that we may with equal justice adopt the simpler expression arrived at in the previous article, and take the effect of an element at S to be according to convenience either $s/\lambda r$, or $s/\lambda p$, where p is the shortest distance from P to the wave-front.

50. Babinet's principle. Two screens may be called complementary when the openings of one correspond exactly to the opaque portions of the other and vice versa. If b be the amplitude at P in the absence of any screen, and a_1, a_2 are vectors representing the vibration at P when either one or the other of two complementary screens is interposed, then the sum of the vectors a_1 and a_2 is obviously equal to b.

The principle due to Babinet allows us, whenever we have calculated the effect of one screen, to obtain that of the complementary screen without further trouble. A little care is necessary in using the principle, to take correct account of the difference in phase. But one simple result may at once be deduced from it. If a_1 is zero, a_2 must be equal to b. Hence at every point where there is no light with one of the screens, the intensity when the complementary screen is introduced, is equal to that observed when the light is unobstructed. This statement cannot however be reversed. If $a_2 = b$, a_1 may have any value between zero and $2a_2$. This is made obvious by the diagram (Fig. 62) in which OA represents the amplitude (b) of the unobstructed light; OB the equal amplitude (a_2) observed when one of the screens is introduced. BA is then that vector which together with OB has OA as resultant. If the point traces out the circle of radius a_2, the vector BA changes in magnitude from zero to $2a_2$.

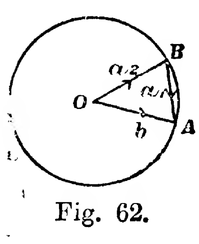


Fig. 62.

51. Shadows of a straight edge in parallel light. Let a plane wave-front *WF* (Fig. 63) fall upon a screen *ME* having a straight vertical edge passing through *E*, the plane of the drawing being horizontal, and let it be required to find the distribution of light on a distant and parallel screen *SS'*. Draw the wave-front which passes through *E*, and divide up that portion *EG* of the wave-front which is not blocked out by the screen, into suitable zones; *EP* being the normal to the wave-front *P* lies on the edge of the geometrical shadow. At *P* the active wave-front *EG* represents one of two exactly symmetrical halves of the complete wave-front, which would operate if the screen were away. Hence the introduction of the screen reduces the amplitude at the geometrical shadow to one half and the intensity to one quarter. To find the amplitude at some point *Q* inside the geometrical shadow, construct Fresnel zones such that

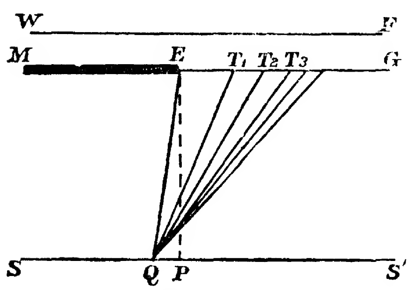


Fig. 63.

$$\frac{\lambda}{2} = T_1Q - EQ = T_2Q - T_1Q = T_3Q - T_2Q = \ldots\ldots$$

Unless *Q* is close to *P*, the resultant vibration due to the different zones will be alternately in opposite directions, and calling the effects of successive zones m_1, m_2, etc. the total effect is

$$m_1 - m_2 + m_3 - m_4 \ldots\ldots$$

In this case the values of *m* diminish too quickly to allow us to write down the sum as $\frac{1}{2}m_1$. It will however be some fraction of m_1, and as with increasing distances of *Q* from *P*, each of the zones diminishes in width, the effect at *Q* is the smaller the further that point lies inside the geometrical shadow. The intensity which as has been shown is only ·25 that of the incident light at the edge of the geometrical shadow, rapidly diminishes still further towards the inside of the shadow and soon becomes inappreciable.

If the point *Q* lies outside the geometrical shadow the intensities are obtained by drawing the normal to the wave-front, and the Fresnel zones, according to Art. 47.

The total effect in amplitude of that portion of the wave-front which lies to the right of the pole, when the shadow-throwing edge is on the left, is equal to ·5, and the effect of the portion included between the pole and the edge is a maximum or a minimum, according as an odd or even number of zones are included between *O* and *E* (Fig. 60). The first maximum takes place when *Q* is at such a distance from *P* that $OE = OT_1$. If the amplitude of the incident light is unity, and the effects of successive zones are m_1, m_2, etc. the first maximum has an amplitude $\cdot 5 + m_1$, half the amplitude of the incident light being added to represent that complete part of the wave which lies to the right of O. When Q has a position such that $OE = OT_2$, there is a minimum with an amplitude $\cdot 5 + m_1 - m_2$. The next maximum has a value $\cdot 5 + m_1 - m_2 + m_3$, and though the maxima and minima rapidly approach each other in magnitude the intensity continues to oscillate about its mean value as the point Q is moved away from the geometrical shadow. The distances (x) of the maxima and minima from the edge are obtained from

$$x^2 = QE^2 - p^2 = \frac{4n-1}{4} p\lambda.$$

The equation shows that the loci of the maxima and minima are parabolas.

TABLE IV.

Shadow of straight edge.

Distance of screen $=100$, $\lambda = 5 \times 10^{-5}$, amplitude of incident light$=1$.

No.	Distance from edge in cms.	Intensities: Outside geometrical shadow	Intensities: Inside geometrical shadow	m	$\sqrt{(4n-1)/2}$
1	·061	1·3748	·0298	1·217	1·225
2	·094	·7774	·0140	1·873	1·871
3	·117	1·1995	·0091	2·345	2·345
4	·137	·8429	·0067	2·739	2·739
5	·154	1·1509	·0053	3·082	3·082
6	·170	·8718	·0044	3·391	3·391
7	·184	1·1259	·0037	3·674	3·674
8	·197	·8891	·0033	3·937	3·937
9	·209	1·1103	·0029	4·183	4·183
10	·221	·9006	·0026	4·416	4·416
11	232	1·0993	·0024	4·637	4·637
12	·242	·9092	·0022	4·848	4·848
13	·252	1·0910	·0020	5·050	5·050

The angle x/p being proportional to $\sqrt{\lambda/p}$ is a small quantity unless n is large. But for large values of n the introduction of the screen causes no appreciable change in the distribution of light. Hence the effect of the screen is confined to the neighbourhood of its geometrical shadow. Table IV. gives the intensities of light at the first seven maxima and six minima outside the geometrical shadow, and the intensities inside at the same distances from the edge. To give an idea of the scale, the positions to which the intensities refer are given for the case in which the shadow is received on a screen one metre away from the linear edge of the shadow-throwing object and the wave-length of light is 5×10^{-5} cms. The meaning of the last two columns will be explained in Art. 52.

The table shows that at a distance of 2·5 mm. from the edge of the geometrical shadow the light inside the shadow has only an intensity equal to the 500th part of that of the incident light, but that outside the shadow, at the same distance, the maximum and minimum intensities still differ by about 20%, while the interval between the bright and dark bands is ·1 mm. The light must of course be homogeneous if it is desired to see more than a few of the bands. The distribution of the intensity of light in the neighbourhood of a straight edge is plotted in Fig. 64 from the numbers given

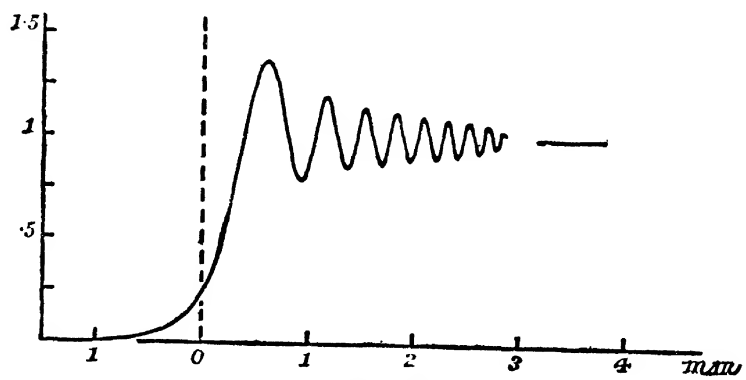


Fig. 64.

by Fresnel. The dotted vertical line represents the edge of the geometrical shadow where the intensity is one quarter. The distance of the screen from the edge is one metre and the scale of abscissæ represents millimetres.

52. Shadow of a straight edge in divergent light. If L (Fig. 65) represents the source of light, which we suppose to be a luminous line parallel to the edge E which throws the shadow, we may for simplicity take the beam to have a cylindrical wave-front with the luminous source as axis. The traces of the wave-front with the plane of the paper are circles. Drawing EG, the wave-front, passing through the edge, we may divide it into laminar Fresnel zones, OT_1, T_1T_2, etc. which satisfy the condition that the

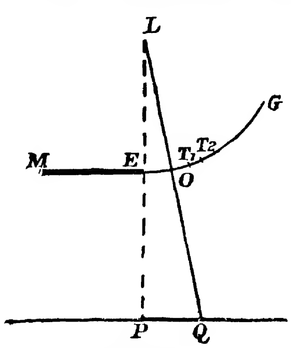


Fig. 65.

resultants of successive zones have opposite phases at Q. The distances of the edges of the laminae must be the same as in the previous article, so that

$$QT_n = QO + \frac{4n-1}{8}\lambda.$$

The condition for the position of the maxima and minima is that a complete number of zones is exposed between O and E so that

$$QE - QO = \frac{4n-1}{8}\lambda.$$

If we put $PQ = x$; $LE = LO = q$,

$$QE = \sqrt{p^2 + x^2} = p + \frac{x^2}{2p} \text{ app.}$$

$$QO = LQ - q = p + \frac{x^2}{2(p+q)} \text{ app.}$$

Hence the positions of the maxima and minima of light are determined by

$$\frac{1}{2}\frac{x^2 q}{p(p+q)} = \frac{4n-1}{8}\lambda,$$

which gives:

$$x = \tfrac{1}{2}\sqrt{p\lambda(4n-1)(p+q)/q}.$$

Fresnel in his celebrated Memoir on Diffraction obtained the expression

$$x = m\sqrt{p\lambda(p+q)/2q},$$

where m is a numerical factor which he calculated by means of the definite integrals which bear his name.

To make our result agree with his, we must put

$$m = \sqrt{(4n-1)/2}.$$

By means of his formula Fresnel obtained an excellent agreement between the observed and calculated positions of the maxima and minima, but the simple method which we have followed gives results which are sufficient for all practical purposes. To show that this is the case, the numerical values of the factor m calculated by Fresnel's method and ours respectively are entered into the two last columns of Table IV. All numbers except the first and second are identical, and even the difference in the position of the first band could hardly be detected by experiment.

As $LQ - QE$ is a constant for a given value of n, it follows that the loci of a maxima and minima are hyperbolas having L and E as foci. The width and hence the effect of each zone may easily be obtained and hence the intensities of the maxima and minima calculated, if desired.

53. Shadow of a narrow lamina. If a cylindrical wave-front WF (Fig. 66) falls on a vertical lamina of which AB represents the horizontal section, and throws a shadow on a screen MN, it is convenient to consider separately the portion of the screen HK which lies within the geometrical shadow and the two other portions which are respectively to the left and right of it. Unless AB is very small, that portion of the wave which passes to the right of B does not affect very considerably the distribution of light to the left of H, and the distribution of light outside the geometrical shadow is therefore approximately that observed outside the shadow of a straight edge bordering a screen of unlimited extent.

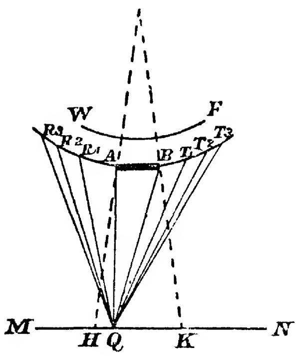


Fig. 66.

To obtain the distribution of light at a point Q inside the geometrical shadow, construct the wave-front passing through A and B and divide it into Fresnel zones. The resultant of the effects of all the zones to the right of B will agree in phase with that due to the first zone, and similarly for the light to the left of A the resultant phase must agree with that of the effect of the first zone. There is a maximum or minimum of light at Q according as the phases resulting from the strips BT_1 and AR_1 act in conjunction or in opposition. Unless Q is very near H or K the first zones may be drawn so that $QT_1 - QB = \frac{\lambda}{2}$ and $QR_1 - QA = \frac{\lambda}{2}$. In that case the first zones act in conjunction or in opposition according as $AQ - BQ$ is an even or odd multiple of half a wave-length. The *positions* of the maxima or minima are therefore the same as if two dependent sources of light were placed at A and B. The space HK is filled in consequence by equidistant bright and dark fringes, but except near the centre of the geometrical shadow the resultant amplitudes of the two portions of the active wave-front are not the same and there is therefore never complete darkness. Near H and K the bands cease to be equidistant and gradually fuse into the ordinary fringes seen outside the shadow. When the lamina is replaced by a thin wire or fibre, the distance between the internal fringes increases, and the position of the external fringes is no longer correctly calculated by considering only one portion of the wave-front. As the width of the obstacle is reduced, the fringes become less distinct and must disappear when the width is only a fraction of a wave-length, for in that case the obstruction is so small that the portions of the wave-front to the right and left of the obstacle cause an amplitude which must be practically identical with that of the unobstructed wave. Plate I. Fig. 5 reproduces a photograph of the shadow of a wire and shows the central bright line.

54. Passage of plane waves through a slit. If a plane wave passes through a slit, placed parallel to the front of a wave, it is easy to obtain an expression for the distribution of light on a distant screen which is parallel to the first. The edges of the slit being supposed vertical, let SS', Fig. 67 and 68, be the intersection of the screen with a horizontal plane and subdivide the slit AB into a large number of vertical strips of equal width. The illumination at a point P is equal to the sum of the effects of the separate strips. If MM' be at a sufficient distance, all parts of the slit produce equal effects as regards magnitude, and the phase difference of the different rays is the same at the screen as on the arc of a circle AK drawn with P as centre. For a distant screen this arc may be taken to be coincident with the line AK drawn at right angles to the direction of the rays (Fig. 67). The phases of the rays proceeding from the centres of successive strips at the points where the rays cross the line AK are in arithmetic progression, and hence if the diagram of vibrations for the point P is constructed, we may apply the results of Art. 5, so that if 2α be the phase difference between the vibrations due to the first and last ray, the resultant vibration has an amplitude $\frac{A \sin \alpha}{\alpha}$ where A is the amplitude at the centre of the pattern. To determine α, we require the phase difference corresponding to the optical distance BK, which if e be the width of the slit and θ the angle between the direction of the rays considered and the normal to the original wave-front is:

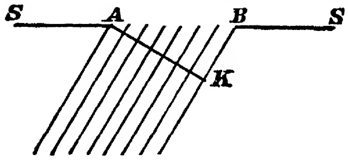


Fig. 67.

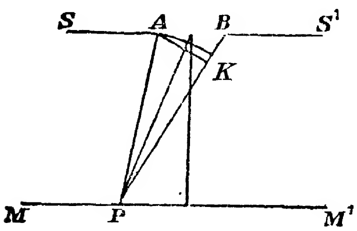


Fig. 68.

$$\alpha = \frac{\pi}{\lambda} e \sin \theta.$$

The illumination at MM' is periodic, the amplitude being zero whenever α is a multiple of π, *i.e.* when $e \sin \theta$ is a multiple of λ. To study the distribution of light more particularly, we must investigate the different values which the function $\left(\frac{\sin \alpha}{\alpha}\right)^2$ takes for different values of α. Its zero values lie at equidistant intervals π. The position of its maxima are found in the usual way from the condition

$$\frac{d}{d\alpha}\left(\frac{\sin \alpha}{\alpha}\right) = 0,$$

which gives
$$\alpha = \tan \alpha.$$

Draw the graph of $\tan\alpha$ as in Fig. 69. The intersections of a straight line drawn at an angle of 45° to the coordinate axes, with the graph, determine the points for which $\tan\alpha = \alpha$. The figure shows that the points of intersection lie on successive branches of the graph and after the first lie near the positions for which α is an odd multiple of a right angle. The first eight values of α for which $\sin\alpha/\alpha$ is a maximum are as follows:

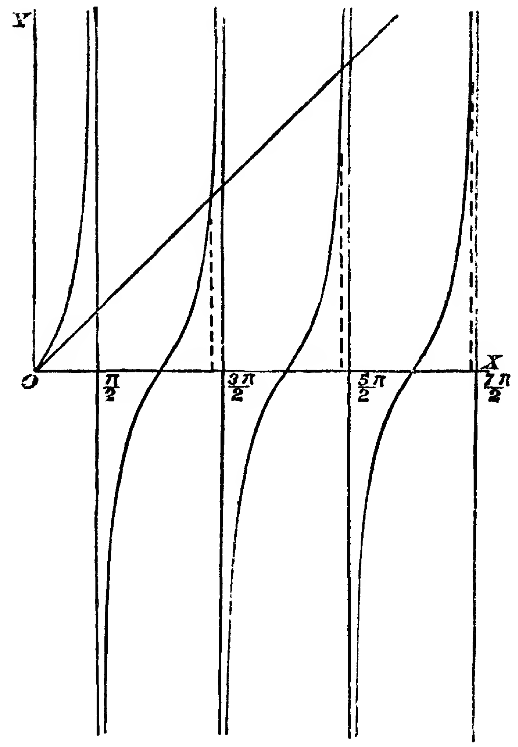


Fig. 69.

$$\alpha_1 = 0,$$
$$\alpha_2 = 1{\cdot}43\ \pi,$$
$$\alpha_3 = 2{\cdot}46\ \pi,$$
$$\alpha_4 = 3{\cdot}47\ \pi,$$
$$\alpha_5 = 4{\cdot}48\ \pi,$$
$$\alpha_6 = 5{\cdot}48\ \pi,$$
$$\alpha_7 = 6{\cdot}48\ \pi,$$
$$\alpha_8 = 7{\cdot}49\ \pi.$$

The curve of amplitudes

$$A = A_0 \sin\alpha/\alpha$$

is drawn in Fig. 70 (dotted line). More important is the intensity curve $I = I_0 (\sin^2\alpha)/\alpha^2$ shown in the same figure. Its coordinates,

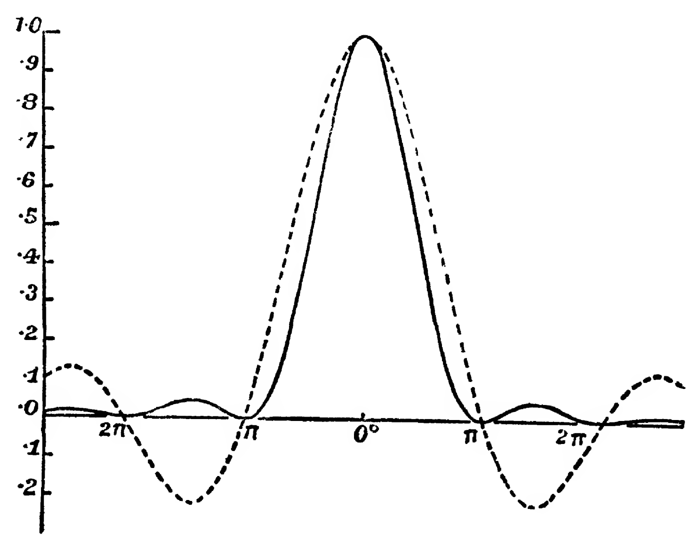


Fig. 70.

when I_0 is equal to one, are given in the third column of Table V. It appears that the bulk of the light is confined to values of α which lie between $\pm\pi$, the intensity of the second maximum being less than $\frac{1}{20}$ of the intensity of that in the central direction. For the first minimum $(\alpha = \pi)$:

$$\sin\theta = \lambda/e.$$

If e is equal to a wave-length, the light spreads out in all directions from the slit, with an intensity which is steadily diminishing as the inclination to the normal increases, but there are no other maxima of light beyond the central one. The equations must in that case be considered as approximate only, as is shown by the fact that the total intensity of light transmitted through the screen would according to the equations be less than the intensity of the light incident on the slit.

TABLE V.

α	$\sin\alpha/\alpha$	$(\sin^2\alpha)/\alpha^2$	α	$\sin\alpha/\alpha$	$(\sin^2\alpha)/\alpha^2$
0°	+1·0000	1·0000	270°	−0·2122	0·04503
15	+0·9886	0·9774	285	−0·1942	0·03771
30	+0·9549	0·9119	300	−0·1654	0·02736
45	+0·9003	0·8105	315	−0·1286	0·01654
60	+0·8270	0·6839	330	−0·0868	0·00754
75	+0·7379	0·5445	345	−0·0430	0·00185
90	+0·6366	0·4053	360	0·0000	0·00000
105	+0·5271	0·2778	375	+0·0395	0·00156
120	+0·4135	0·1710	390	+0·0735	0·00540
135	+0·3001	0·0901	405	+0·1000	0·01001
150	+0·1910	0·0365	420	+0·1181	0·01396
165	+0·0899	0·0081	435	+0·1272	0·01619
180	0·0000	0·0000	450	+0·1273	0·01621
195	−0·0760	0·00578	465	+0·1190	0·01416
210	−0·1364	0·01861	480	+0·1034	0·01069
225	−0·1801	0·03242	495	+0·0818	0·00670
240	−0·2067	0·04274	510	+0·0562	0·00315
255	−0·2170	0·04710	525	+0·0282	0·00080
			540	0·0000	0·00000

For values of e smaller than λ, the equations must ***a fortiori*** not be taken as giving more than an approximate representation of the facts, which may be wide of the truth if e is a small fraction of the wave-length.

When e is large compared with the wave-length, the whole light is confined to directions for which θ is very small. This explains the apparent discrepancy between the behaviour of sound and light, which retarded so long the general adoption of the undulatory theory of light. The amount of the spreading of waves which have passed through an opening depends entirely on the relation between the wave-length and the opening. If sound-waves, having a length measured in feet, pass through an opening, the linear dimensions of which are of about the same magnitude, the waves expand in all directions, but if light-waves pass through the same openings, the spreading is practically nil, owing to the fact that the length of the waves is now very minute in comparison with the opening, and hence there is destruction of light by interference in oblique directions. To make experiments of sound and light waves comparable with each other, the openings should be made proportional to the lengths of the waves.

55. Passage of light through slit. General case. In the previous article it has been assumed that the screen receiving the light is at a great distance. We may now consider the more general case in which the screen is nearer and the incident light divergent. If Fig. 71 represents a horizontal section, L being the linear source and AB the aperture, we may find the amplitude at a point Q of the screen MM' by dividing the wave-front between A and B into appropriate zones. Consider first the light at the central point P. If O be the central point of the wave-front between A and B and the screen be at such a distance that $PA - PO = (4n-1)\lambda/8$ each half OA and OB of the wave-front contains an even or odd number of zones according as n is even or odd. Hence there is a maximum or minimum of light at P according as n is odd or even. As the screen is brought nearer, the observed system of fringes will alternately have a bright or dark centre at P. If p and q be the distances of P and L from the plane of the aperture, and d half the aperture of AB,

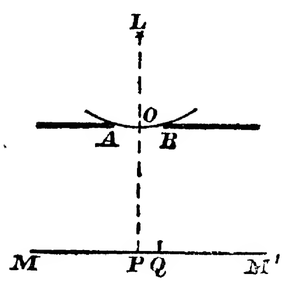


Fig. 71.

$$PO = p + q - \sqrt{q^2 + d^2}$$

$$= p - \frac{d^2}{2q} \text{ app.}$$

$$PA = \sqrt{p^2 + d^2}$$

$$= p + \frac{d^2}{2p} \text{ app.}$$

$$\therefore\ PA - PO = \frac{d^2}{2}\left\{\frac{1}{p} + \frac{1}{q}\right\}$$

and therefore $$\frac{1}{p} + \frac{1}{q} = \frac{4n-1}{4}\,\frac{\lambda}{d^2}$$

determines the distance p of the screen from the opening, the central fringe being bright when n is odd and dark when n is even. When the point Q is not included in the geometrical beam of light which is bounded by the straight lines LB and LA, a similar reasoning leads to the conclusion that there is the centre of a bright or dark fringe at Q according as $AQ-BQ$ is an odd or even multiple of half a wave-length.

56. Passage of light through a circular aperture. When the perforations in a screen are such that we can divide the screen into circular zones, the calculation of the intensities is very simple for points in the axis of the zones.

Let O (Fig. 72) be the centre of a small circular aperture in a screen, and OP a line at right angles to the screen which we shall call the axis. If it is required to determine the amplitude at P due to a wave-front of unit amplitude incident on the screen, which we shall consider in the first instance to be plane and parallel to it, we may divide the aperture into Fresnel's zones, which produce effects which are equal in magnitude but alternately opposite in direction. If the radius OR of the aperture is such that an *even* number of zones is included, the amplitude at P is zero; if an *uneven* number is included the amplitude is a maximum and equal to that due to the first zone, and therefore double that of the unobstructed wave. The introduction of the screen with small aperture doubles the amplitude therefore at certain points. The condition for maximum or minimum of light is if $PO=p$, $OR=r$,

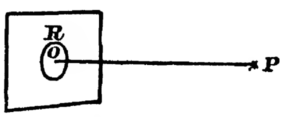


Fig. 72.

$$\frac{n\lambda}{2}=\sqrt{p^2+r^2}-p=\frac{r^2}{2p}\text{ app.}$$

where there is a maximum if n be odd and a minimum if n be even. The general expression for the amplitude on the axis is found by subdividing the aperture into a large number of small zones of equal areas. Their total effect, according to Art. 5, is $(A\sin\alpha)/\alpha$ where for α we must put half the difference in phase at P of the disturbances due respectively to the first and last zone, *i.e.* half the difference in phase corresponding to an optical length $\frac{1}{2}n\lambda$. This gives:

$$\alpha=\frac{\pi}{\lambda}\cdot\frac{n\lambda}{2}=\frac{\pi r^2}{2p\lambda}.$$

A is the amplitude at P calculated on the supposition that the disturbances of all zones reach P in the same phase, which would according to Art. 46 be $\pi r^2/p\lambda$, *i.e.* the area of the aperture divided by $p\lambda$.

The amplitude at P is therefore $2\sin(\pi r^2/2p\lambda)$. The points of zero illumination which have already been determined are the nearer together the smaller the distance of P. Sideways from the axis, the amplitudes cannot be calculated by simple methods, but general considerations similar to those which lead to accurate results in the case of long rectangular openings, are sufficient to show that there must be rhythmical alternations in the illumination. Hence a screen placed across the axis will show bright and dark rings having at P a bright or dark centre according to the distance of P from the opening.

The case of a divergent beam of light presents no further difficulty. We may subdivide the spherical wave-front into zones of equal area and obtain again at P the amplitude $\dfrac{A\sin\alpha}{\alpha}$ with the difference that $\alpha = \dfrac{\pi}{\lambda}r^2\left(\dfrac{1}{2p}+\dfrac{1}{2q}\right)$, q being the distance of L from the screen. A has the same value as before. Hence the points of maximum and minimum illumination are determined by

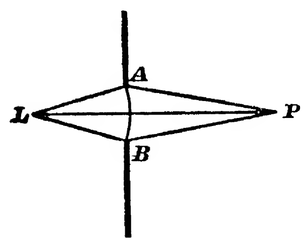


Fig. 73.

$$\frac{1}{p}+\frac{1}{q}=\frac{n\lambda}{r^2},$$

and the amplitude at the maximum is $2q/(p+q)$.

57. Shadow of a circular disc. OR (Fig. 74) being a circular disc, a spherical wave-front diverging from L, a luminous point on the axis of the disc, will throw a shadow on a screen SS', the centre of the shadow being on the axis. If Fresnel zones are drawn on the wave-front, the total effect at P as regards amplitude may be determined as in Art. 46 to be the same as that due to half the first zone, and if the disc is small, the first zone surrounding the edge of the disc has the same area as the central zone at O, which is covered by the disc. Hence the illumination at P is the same as if the disc were away. Round this central bright spot there are alternately dark and bright rings. It will be an interesting exercise for the student to deduce the constancy of illumination on the axis of a shadow-throwing disc from Babinet's principle, making use of the amplitude at the bright and dark centres of the complementary circular aperture. The fact that the shadow of a circular disc has a bright spot at its centre was discovered experimentally in the early part of the 18th century, but had been forgotten again when about 100 years later Poisson deduced it as a consequence of the wave-theory of

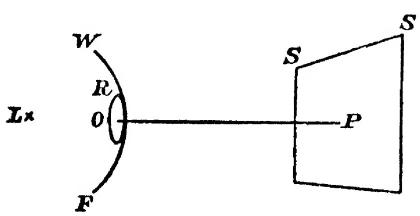


Fig. 74.

light. Arago, who was unaware of the earlier experiment, tested Poisson's mathematical conclusion, and verified it.

58. Zone plates. On a plane screen draw with O as centre, circles which divide the Fresnel zones with respect to a point P on the normal OP, the wave-front being supposed to be plane. For the radii of the circle we have the relation.

$$r^2 = np_0\lambda,$$

where p_0 is the distance OP, and where n takes the values 1, 2, 3 etc. for successive circles. Imagine the zones on the screen to be alternately opaque and transparent. Then if a wave-front proceeding in the direction PO falls on the screen, the phases due to all transparent zones are in agreement at P, and hence the amplitude at P will be $\frac{1}{2}Nm$ where m represents the effect of the first zone and N the total number of zones.

The amplitude at P will therefore be N times what it would be if the screen were away. Such a zone plate acts like a lens concentrating parallel light to a focus, the focal distance being p_0. If now the source of light is moved to a point q from the screen, the zones will again unite their effects at P provided (Art. 56)

$$\frac{1}{p} + \frac{1}{q} = \frac{n\lambda}{r^2},$$

$$\therefore \frac{1}{p} + \frac{1}{q} = \frac{1}{p_0}.$$

The relation between object and image is therefore the same as for a lens.

Zone plates may be made by drawing circles on a sheet of paper, the radii of which are as the square roots of successive numbers, and painting the alternate zones in black. When a photograph on glass is taken of such a drawing, a plate is produced which satisfies the conditions of a zone plate. To prepare an effective zone plate involves great labour. Prof. R. W. Wood has published a reduced print of such a plate* from which other still more reduced copies may be prepared by photographic reproduction. Prof. Wood† has also described a photographic method by means of which zone plates may be made, which give for alternate zones a complete phase reversal. A more perfect imitation of a lens may thus be obtained.

59. Historical. Augustin Jean Fresnel was born on May 10th, 1788, in Normandy, and entered the Government service as an engineer. He was occupied with the construction of roads, but lost his position owing to his having joined a body of men who opposed

* *Phil. Mag.* XLV. p. 511. 1898. † *Ibidem.*

Napoleon's re-entry into France, after his escape from Elba. Reinstated after Waterloo, he remained some time living in a small village in Normandy where his first study of the phenomena of diffraction seems to have been made. Fresnel was always of weak health and died on July 14, 1827. The undulatory theory of Optics owes to Fresnel more than to any other single man. His earlier work on Interference had to a great extent been anticipated by Thomas Young, but he is undoubtedly the discoverer of the true explanation of Diffraction. Young had tried to explain the external fringes of a shadow by means of interference of the rays which passed near the shadow-throwing object and those that were reflected from its surface. Fresnel, starting with the same idea, soon found that it was wrong, and proved by conclusive experiments that the surface reflexion had nothing to do with the appearance of the fringes. He then showed by mathematical calculation that the limitation of the beam, by the shadow-throwing object, was alone sufficient to cause the rhythmic variations of intensity outside the shadow.

CHAPTER VI.

MEASUREMENT OF WAVE-LENGTHS.

60. General theory of a grating. A grating is a surface having a periodical structure which impresses a periodical alteration of phase or intensity on a transmitted or reflected wave of light. The most common method of manufacturing a grating is to rule equidistant lines with a diamond point on a surface of glass or metal. The diamond introduces a periodical structure, each portion of which is probably very irregular, but which is repeated at perfectly regular intervals, Fig. 75. If the grating is ruled on a plane surface, that surface is called the plane of the grating. Any plane passing through corresponding points of the grooves such as A_1, A_2, A_3, is parallel to the plane of the grating. We distinguish between "reflexion gratings" and "transmission gratings" according as they are ruled on an opaque surface, the reflected or scattered light being used, or a transparent plate, through which the light is transmitted.

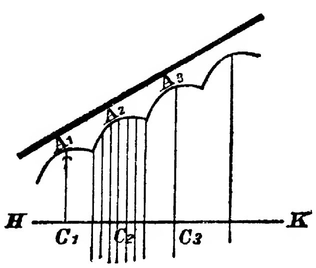


Fig. 75.

Let a plane wave-front be incident parallel to the grating. Waves spread out from the different portions of the grooves which may be considered as centres of secondary disturbances. If the light be received on a distant screen, the resultant of all vibrations at each point may be determined. Consider that point of the screen which lies in a direction A_2C_2 from the grating, and draw a plane HK at right angles to that direction. As the optical distance from any point on HK to the corresponding point of the distant screen is the same, the phase differences between individual secondary rays in any one direction are the same at the screen as they are at HK. Also we may treat the rays which are ultimately brought together as having equal amplitudes, the small differences in the distances from different parts of the grating to the screen being negligible. We combine in the first place, those vibrations which are due to the secondary waves coming from one of the grooves. Selecting any point on the groove A_2, we may always express the phase of the resultant vibration due to the whole groove as that corresponding to an optical distance $A_2C_2 - \epsilon$, where ϵ is some length which depends on the shape of the groove and on the direction of A_2C_2. The resultant amplitude similarly may be written ka, where a is the amplitude of the incident light and k a factor depending also on the shape of the grooves and the direction. Taking the resultant of the other grooves, we should find similarly that the resultant phases at HK may be derived from the optical distance $A_1C_1 - \epsilon$, $A_3C_3 - \epsilon$, etc., A_1, A_2, A_3, being corresponding points on the grating. The theory of the grating depends on the fact that the values of ϵ and k are the same for each groove. This involves the similarity of all the grooves, and if that similarity holds, the difference in phases between the resultant vibrations of two successive grooves is $(A_2C_2 - \epsilon) - (A_1C_1 - \epsilon)$ and is therefore independent of ϵ. We may now draw a plane through any set of corresponding points of the groove and call it the plane of the grating (Fig. 76), and in calculating the resultant phases at HK we need only consider the difference in the optical distance A_1C_1, A_2C_2, A_3C_3....... If that difference is a multiple of a wave-length, the phases at HK are identical and we must then obviously have a maximum of light, wherever those identical phases are brought together. This may either be the distant screen or the principal focus of a lens placed with its axis at right angles to HK. The direction in which these maxima appear is easily obtained. If θ be the angle between the normal to the grating and the direction A_1C_1 and A_1N be drawn at right angles to A_1C_1:

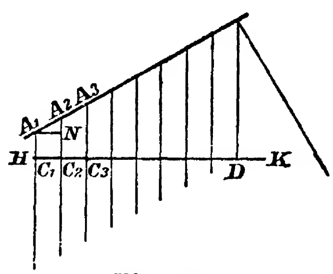


Fig. 76.

$$\sin\theta = \frac{A_2N}{A_1A_2} = \frac{n\lambda}{e} \text{ .... (1),}$$

where e is the distance A_1A_2 between the grooves ruled on the grating, λ the wave-length and n an integer number. The number of maxima is finite because $\sin\theta$ cannot be greater than one, and the highest value which we can take is therefore that integer which is nearest but smaller than e/λ. If e were smaller than λ there could be no maximum except that for which $n = 0$. The amplitudes in the direction of the maxima are Nka, where N is the total number of grooves and k the constant already introduced, which may and does very seriously affect the amplitude. It is theoretically possible that k is zero for one of the directions defined by (1) and in that case that maximum would of course be absent. It is also possible that k is unity, and in that case the whole of the light would be concentrated at or near that maximum.

The complete investigation of the grating includes the determination of the amplitudes of light in directions not necessarily confined to those at which the maxima appear. We proceed, therefore, to find the distribution of light in the neighbourhood of the maxima. The wave-length of a homogeneous beam incident on the grating being λ and having, as has been shown, a maximum in such directions that (Fig. 76) $A_2N = n\lambda$, let the whole system of rays A_1C_1, A_2C_2 etc. and with it the normal plane HK be turned round slightly so that A_2N now becomes $n\lambda'$, where λ' is a length differing little from λ. The difference in phase between the vibrations at C_2 and C_1 for the wave-length λ becomes $2\pi n\lambda'/\lambda$ or $2\pi n(\lambda' - \lambda)/\lambda$, as we may add or subtract any multiple of four right angles to a phase difference. This is also the phase difference between the vibrations at C_3 and C_2, etc. To obtain the complete resultant, we can therefore apply the proposition of Art. 5, which gives for the amplitude of N vibrations of equal amplitude ka, and constant phase difference $2\alpha/N$, a resultant amplitude

$$Nka\frac{\sin\alpha}{\alpha}.$$

In the present case, $\alpha = \pi nN(\lambda' - \lambda)/\lambda$.

The distribution of intensity corresponding to this amplitude has been discussed in Art. 53. Fig. 70 shows for different values of α, the amplitude $(\sin\alpha)/\alpha$ (dotted curve) and the intensity $(\sin^2\alpha)/\alpha^2$ (full curve). The intensity has secondary maxima which are not, however, important compared with the principal one, at which $\alpha = 0$.

The amount of light is everywhere small when α is greater than 2π; hence if Nn is large, the light is concentrated nearly in those directions for which $(\lambda' - \lambda)/\lambda$ is very small. It is owing to the rapid falling off of the light from both sides of the principal maxima, that the grating can be made use of to separate the different components of white light, and to produce quasi-homogeneous vibrations.

The condition for the first minimum $\alpha = \pi$, leads to

$$\lambda'/(\lambda - \lambda') = nN \qquad (2).$$

It will be shown in Chapter VII. that a spectroscope resolves a double line, the components of which have wave-lengths λ and λ', when the maximum of the diffraction image of one line coincides with the first minimum of the other. The greater the value of Nn the smaller is the difference $\lambda - \lambda'$ which may be resolved. We may therefore take nN to be a measure of the resolving power.

The incident wave-front has so far been taken as parallel to the plane of the grating. For oblique incidence, consider a grating formed by ruling lines on a glass surface, and let a plane wave be transmitted obliquely through it. Let A_1, A_2 (Fig. 77) be corresponding points on successive grooves, and LM the incident wave-front, inclined at an angle ϕ to the plane of the grating. Draw two rays LA_1, MA_2, and consider the light diffracted in the direction A_1C_1, inclined at an angle θ to the normal of the grating. Draw A_1N and A_2T at right angles to A_1C_1 and A_1L respectively. The difference in phase between C_1 and C_2 is then

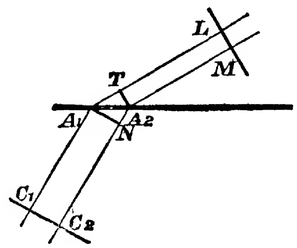


Fig. 77.

$$e(\sin\phi - \sin\theta),$$

and there is a maximum when

$$e(\sin\phi - \sin\theta) = \pm n\lambda \quad\text{.......................(3).}$$

θ and ϕ are here taken as having the same sign when they are both on opposite sides of the normal.

Writing γ for $\phi - \theta$, the angle between the incident and diffracted beams, the condition for a minimum or maximum of deviation is $\frac{d\gamma}{d\theta}=0$, which leads to $d\phi = d\theta$. By differentiating (3) we obtain

$$\cos\phi d\phi - \cos\theta d\theta = 0.$$

If $d\phi = d\theta$ it follows that $\cos\phi = \cos\theta$, *i.e.* $\phi = \pm\theta$. ϕ and θ cannot be equal unless $n = 0$, which case need not be considered. For the condition of maximum-minimum we have therefore $\phi = -\theta$, which shows that the incident and diffracted light form equal angles with the plane of the grating. Further consideration shows that it is a minimum and not a maximum deviation that is involved.

If $\phi = -\theta$ the deviation is 2θ. Equation (3) becomes in that case

$$2e\sin\frac{\gamma}{2} = 2e\sin\theta = n\lambda.$$

61. Overlapping of spectra. The maxima of light for normal incidence have been shown to take place when $e\sin\theta = n\lambda$. For each value of n, the maxima of the different wave-lengths take place along different directions, and hence the grating "analyses" the light falling on it and produces quasi-homogeneous light. It acts in this respect like a prism, but splits up the light into a number of spectra, each value of n giving a separate spectrum. For $n = 0$, there is a maximum, but there is no spectrum because the position of the maximum is independent of the wave-length. The direction of this maximum is the direction of the incident light in a transmission grating, while in a grating which acts by reflexion it is the direction in which the incident beam would be reflected from a polished surface coincident with the grating. For $n=1$, we have the so-called spectrum of the first order, which spreads over the quadrant between $\theta=0$ for $\lambda=0$ and $\theta=\frac{1}{2}\pi$ for $\lambda=e$. Similarly the spectrum of the second order, for which $n=2$, spreads over the same quadrant, the limits of wave-length being $\lambda=0$ for $\theta=0$, and $\lambda=e/2$ for $\theta=\frac{1}{2}\pi$. For each value of θ we have therefore an infinite number of overlapping maxima corresponding to all wave-lengths which are submultiples of $e\sin\theta$. If we confine ourselves to eye-observations, we need only consider the wave-lengths lying between 4×10^{-5} and 8×10^{-5}. The limits θ' and θ of the spectra of different orders are then with normal incidence

$$\begin{aligned}
&\text{for } n=1; \quad 4\times10^{-5}=e\sin\theta' \text{ and } 8\times10^{-5}=e\sin\theta,\\
&\text{for } n=2; \quad 8\times10^{-5}=e\sin\theta' \text{ and } 16\times10^{-5}=e\sin\theta,\\
&\text{for } n=3; \quad 12\times10^{-5}=e\sin\theta' \text{ and } 24\times10^{-5}=e\sin\theta,\\
&\text{for } n=4; \quad 16\times10^{-5}=e\sin\theta' \text{ and } 32\times10^{-5}=e\sin\theta.
\end{aligned}$$

In Fig. 78 the extension of the different spectra is marked by straight lines lying above each other to avoid actual overlapping. The wave-lengths marked are those corresponding to the first order spectrum; the wave-lengths belonging to the spectrum of order n are obtained by dividing these numbers by n.

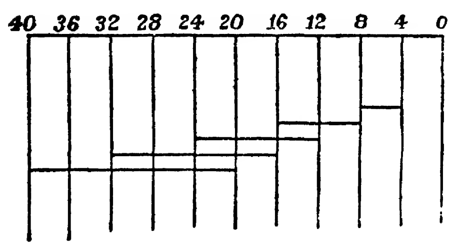


Fig. 78.

The visible spectrum of the first order stands out clear of the rest; but the second and third overlap to a great extent, the range between $\lambda=6\times10^{-5}$ and $\lambda=8\times10^{-5}$ of the second order being coincident with the range of $\lambda=4\times10^{-5}$ to $\lambda=5{\cdot}3\times10^{-5}$ of the third order. The spectra of higher orders spreading over greater ranges of θ overlap more and more, and special devices have to be adopted to separate the spectra, when observations are made in the higher orders. When spectra are to be recorded by photography, there is a similar overlapping but its range is different.

62. Dispersion of gratings. The maxima of two wave-lengths λ_1 and λ_2 being in such positions that

$$e\sin\theta_1=n\lambda_1,$$
$$e\sin\theta_2=n\lambda_2,$$

the ratio $(\theta_1-\theta_2)/(\lambda_1-\lambda_2)$ may be taken to measure the angular dispersion of the grating. The ratio increases with increasing values of $n\lambda$ and hence the dispersion increases with the order of the spectrum.

If the incident beam is oblique

$$e(\sin\theta - \sin\phi) = n\lambda,$$

which, by differentiation, gives with a constant value of ϕ

$$e\cos\theta d\theta = n d\lambda,$$

so that the angular dispersion is $\dfrac{d\theta}{d\lambda} = n/e\cos\theta$.

For a given order the dispersion is therefore inversely proportional to $\cos\theta$ and involves ϕ and λ only in so far as θ depends on these quantities.

When the diffracted beam leaves the grating nearly normally, $\cos\theta$ varies very slowly. In that case the dispersion is proportional to the order of the spectrum and independent of the wave-length, *i.e.* equal angular separation means equal differences of wave-length. We then say that the spectrum formed is "normal."

63. Resolving power of gratings. The use of a grating as an analyser of light depends on its power to form a pure spectrum. To obtain a measure of the purity of a spectrum, we may imagine it to be projected on a screen, which has a narrow opening parallel to the original slit intended to transmit only that wave-length which has a maximum coinciding in position with the opening. It is then found

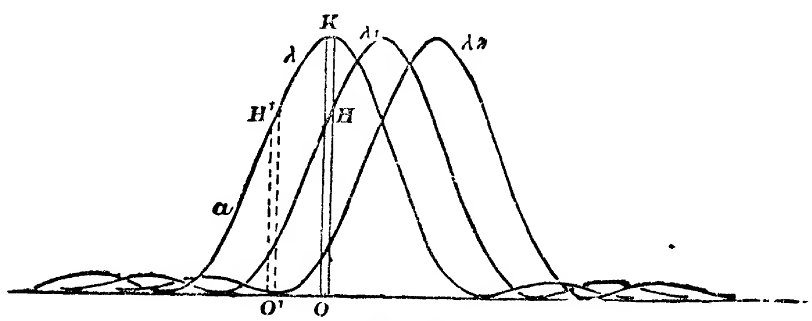


Fig. 79.

that the waves passing through even an indefinitely narrow aperture are not absolutely homogeneous. In Fig. 79 the curve a represents the distribution of light on the screen for a given wave-length. OK indicates the position of a narrow opening placed so as to transmit the maximum amount of light having a given wave-length λ, the amount so transmitted being proportional to the intensity OK and to the width of the opening. If λ_1 be a wave-length near λ, it will have its maximum a little to one side. Its intensity curve is represented by the second curve and an amount of its light proportional to OH passes through the opening. The curves of intensity having no definite limit, there is some light of every wave-length passing through the slit, but the intensity quickly diminishes and we need only consider those wave-lengths which are not very different from λ. If we wish to compare different spectrum-forming instruments with each other, it will be sufficient to limit the investigation to that light which lies between the two minima on either side of the maximum.

It follows from Art. 60 that a wave-length λ_1 has its first minimum where there is maximum for λ if $nN(\lambda_1-\lambda)/\lambda=\pm 1$. Hence we may say that the range of wave-lengths passing through the opening extend from a wave-length $\lambda\left(1+\frac{1}{Nn}\right)$ to a wave-length $\lambda\left(1-\frac{1}{Nn}\right)$. The quantity Nn has been called the resolving power of the grating. Denoting it by R, we may say that very little light passes through the slit which differs in wave-length from λ by more than λ/R. Resolving power will be further considered in Chapter VII.

64. Action of grating on impulses. In the discussion of the grating its action on homogeneous vibrations have so far been made the starting point, but a clearer view is obtained by imagining the disturbance to be confined to an impulsive velocity spread equally over a plane wave-front. Such an impulse, as we have already seen, represents white light, and by treating such light as an impulse we gain the advantage of having to consider a single entity in place of an infinite number of overlapping waves of infinite extent. We shall also be led to an instructive representation of homogeneous light based on white light. Without wishing to give to one of these views the preference over the other, we must emphasize the justification of both, believing that a clear idea of the phenomena of light can only be obtained by a proper recognition of the duality of the relationship between white and homogeneous light.

In Fig. 76, Art. 60, let the incident light consist of a single impulse spread over a plane wave-front which is parallel to the grating. The impulsive motion will reach the points C_1, C_2, C_3, at regular intervals. If therefore a lens be placed in such a position that a wave-front HK would be brought together at its principal focus, a succession of impulses would pass that focus at regular intervals of time, the result being a periodic disturbance.

There will be as many impulses as there are lines on the grating and the interval between them is equal to the time which the disturbance takes to travel through the distance $e \sin\theta$. The whole theory of the grating is contained in this statement. It would be easy to show that the overlapping of spectra, and the partial homogeneity which becomes more and more perfect as the number of lines on the grating is increased, are all implied in the finite succession of impulses and it might be instructive to do so, but there is no necessity for it. The sole object of Physics is to explain what we can observe, and we should turn our attention therefore to the physical phenomena which the light after reflexion from a grating exhibits. For this purpose the impulse serves at least as well as the homogeneous radiation. We should enquire therefore what are the effects of such a finite succession of impulses on our eye, on a photographic plate or an absorbing medium. In each of these cases resonance plays the predominant part, and our problem resolves itself therefore into finding the resonance effects which may be caused by a succession of impulses and to compare them—if we wish—with those of homogeneous vibrations.

The analogy of sound may help us. If a blast of air be directed against a rotating disc perforated at regular intervals like the disc of a siren, a musical sound is heard; or to make the analogy with the grating more complete, imagine a sharp noise of very short duration to be reflected from a railing, when the reflected impulses returning at regular intervals may produce the effect of a musical note. In order to examine the resonance effects which a succession of impulses is capable of producing, we take the case of a pendulum set into a motion by a blow succeeded by others at regular intervals. If τ is the period of the pendulum, τ' that of the interval between the blows assumed to be slightly greater than τ, the second blow will be delivered when the pendulum has just passed the position of equilibrium and will have practically the same effect in increasing the momentum as the first; the same is the case for the succeeding blows which will all increase the swing of the pendulum until the accumulated difference in period is such that the forward blows are delivered when the pendulum swings backwards.

The difference between τ and τ' therefore becomes serious when $N(\tau'-\tau)=\frac{1}{4}\tau$, N being the number of blows delivered. If the difference between τ' and τ is less than that indicated by the equation, we should be unable to distinguish between the time interval of the blows and the period of the pendulum, and if we were to investigate the succession N of impulses by some resonance method, we should be driven to the conclusion that it contained all periodicities between the limits $\tau\left(1 \pm \frac{1}{4N}\right)$ in almost equal proportion. Outside these limits there is still some resonance but with diminishing effect. It is seen that the greater the number N the more nearly can we identify the disturbance with a homogeneous vibration. In the case of sound the matter may perhaps be put somewhat clearer by superposing the succession of impulses on a periodic homogeneous vibration and examining the "beats" produced. If $N\tau'=(N\pm 1)\tau$ the note has been alternately increased and weakened, and the ear would, by the alteration in intensity, clearly perceive that it is dealing with disturbances of different periods. But if $N\tau'$ lies anywhere between the limits $(N\pm\frac{1}{4})\tau$, there will be little variation in intensity and the ear could not form any definite conclusion as to any difference between τ' and τ. We should conclude that the sound examined contained all the periods included within the narrower limits given in about equal proportion, but that in agreement with previous results, it is only when $N\tau'$ lies outside $(N \pm 1)\tau$ that we can altogether neglect the periodicity. The quasi-homogeneous effect of a succession of impulses and its approach to homogeneity as their number increases is thus explained. There is a peculiarity of the periodicity produced by the succession of impulses inasmuch as it is impossible to distinguish between the periodicity τ and the periodicity $\frac{1}{2}\tau$, $\frac{1}{3}\tau$, or $\frac{1}{n}\tau$: which are all equally contained in it. A consideration of the resonance effect shows that the succession of blows has the same effect whether the pendulum in the meantime has performed one, two, or n complete oscillations. This explains the overlapping spectra in a grating. We have used the effects of resonance to pick out the periods contained in a succession of impulses such as is formed by a grating, but the mathematician will not find it difficult to apply Fourier's analysis and to express directly the impulses in a series proceeding by sines and cosines. He may thus easily convince himself that our representation of the effects of the grating is in all respects identical whether the white light is decomposed into homogeneous vibrations at its source or after it emerges from the grating.

65. Talbot's bands. If, while the spectrum formed by a prism or grating is observed, half the pupil of the eye be covered with a thin plate of mica or glass, the spectrum is seen to be traversed by dark bands, provided the plate is inserted on that side on which the blue of the spectrum appears. These bands were first observed by Fox Talbot. Instead of viewing the spectrum directly we may use a telescope, the plate being inserted on the side of the thin edge of the prism forming the spectrum, so as to cover a portion of the aperture of the object glass.

Similar bands were observed by Baden Powell, who used a hollow glass prism with its refracting edge downwards and filled with some highly refractive liquid, into which he inserted a plate of glass from above so that its plane approximately bisected the angle of the prism. The plate was only pushed half way down the liquid, so as to leave its lower parts clear. Interference bands then appeared, but only when the refractive index of the liquid was greater than that of the glass. Stokes subsequently showed that when the refractive index of the glass was the greater of the two, the bands could still be observed; only in this case it was necessary to place the plate in the thinner end of the prism, leaving the thicker part clear.

A simple explanation of these bands is sometimes based on the consideration that the two portions of the light, which, in the absence of the interposed plate, would reach the retina in the same phase, are retarded relatively to each other by the plate, so that interference may take place. This reasoning is obviously wrong, for it does not explain the manner in which the effects depend on the relative refractive index of the liquid and the inserted plate.

A more complete explanation taking account of this asymmetry has been given by Airy and Stokes, and involves an elaborate mathematical process. A very simple treatment may be given if, instead of basing the calculation on Fourier's analysis, we consider the source of the light to emit a succession of impulsive velocities. In Fig. 76 (Art. 60) we may consider the wave-front to consist of a simple impulse which reaches the grating so that the points A_1, A_2, A_3, etc. are simultaneously disturbed. At the plane HK, the disturbance will reach the points C_1, C_2, C_3, in succession, and if a lens be placed with its axis at right angles to HK, a disturbance will pass the focus of the lens at regular intervals of time, as already explained.

The question now is: How can the impulses which succeed each other at the focus of the lens be made to interfere with each other? Clearly only by retarding those which reach the focus first or by accelerating those which reach it last. A plate of appropriate thickness introduced on the left-hand side of the figure as it is drawn could be made to answer the purpose. If, on the contrary, the same plate be introduced on the right-hand side, it would only retard those impulses which already arrive late, and therefore no interference could take place. If the retardation be such that the retarded impulses fall just half-way between the original impulses, interference is complete because the primary periodicity is destroyed. There is however an increased amplitude of the half period and its submultiples. We find therefore a destruction of the spectra of odd orders and an increased illumination in the spectra of even orders.

There is one thickness of the plate for which the bands are seen most sharply. This is clearly the thickness which gives a retardation such that there is most complete overlapping, and hence we see that the retardation must be such that the retarded impulse coming from the first line of the grating, and the unretarded impulse coming from the central line, arrive together. This means that the retardation is $\frac{1}{2}N\lambda$, if N is the total number of lines in the grating the plate must be pushed sufficiently far into the beam to affect half its width. The wave-length λ here means the wave-length of that homogeneous train of waves which has its first principal maximum at the focus of the lens. If the retardation has more than twice its most effective value, the series of impulses from the first half of the grating pass through the focus later than those coming from the second half, and hence interference ceases.

If at a certain point of the spectrum corresponding to a wave-length λ there is a maximum of light, the relative retardation of the two interfering impulses must be equal to $m\lambda$, m being an integer; the next adjoining band towards the violet will appear at a wave-length λ' such that $m\lambda = (m+1)\lambda'$.

Hence for the distance between the bands

$$\frac{\lambda - \lambda'}{\lambda} = \frac{1}{m},$$

with the best thickness of interposed plate, $m = \frac{1}{2}N$, and hence $(\lambda - \lambda')/\lambda' = 2/N$ where λ' in the denominator may with sufficient accuracy be replaced by λ.

If a linear homogeneous source of light of wave-length λ be examined by means of a grating, the central image extends to a wave-length λ_1 such that

$$\frac{\lambda - \lambda_1}{\lambda} = \frac{1}{N},$$

where N, as before, is the total number of lines on the grating.

Hence the following proposition:—If, in observing Talbot's bands, the best thickness of retarding plate be chosen, the distance between each maximum and the nearest minimum is equal to the distance between the central maximum and the first minimum of the diffractive image of homogeneous light, observed in the same region of the spectrum with the same optical arrangement.

If we use prisms instead of a grating, the number of lines N must be replaced by the quantity which corresponds to it as regards resolving power, viz., $td\mu/d\lambda$ where t is the aggregate effective thickness of the prisms. It follows that the retardation which gives the best interference bands with prisms is $\frac{1}{2}\lambda t d\mu/d\lambda$*.

66. Wire gratings. In certain cases, the intensity of the spectra of different orders may be calculated. If the grating is formed by a number of equidistant thin wires of equal thickness (Fig. 80), the periodicity of the grating is such that one portion does not obstruct the passage of the light whilst the other is opaque. Take the incident light to be normal to the grating, and let the widths of each transparent and opaque portion be a and b respectively; the amplitude of the light diffracted at an angle θ to the normal is then (Art. 54) $(A \sin \alpha)/\alpha$ where $\alpha = \pi a \sin\theta/\lambda$.

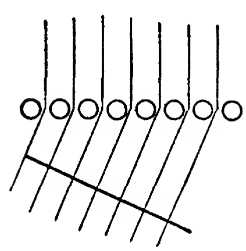


Fig. 80.

The maximum of the nth order is determined by

$$(a+b)\sin\theta = n\lambda; \text{ so that } \alpha = \pi a n/(a+b).$$

The amplitudes at the maxima are therefore

$$\frac{A(a+b)}{\pi n a} \sin \frac{n\pi a}{a+b}.$$

* *Phil. Mag.* Vol. VII. p. 1 (1904).

For the central image, in which there is no dispersion, $a = 0$ and the amplitude is A. The law of falling off in the intensities at the sides of the maximum in each diffraction image is the same for all maxima, so that for the ratio of the intensities of the images, we may substitute the ratio of the squares of the amplitudes at the maxima. For the calculation of the amplitude at the central maximum, it is sufficient to point out that the interposition of the grating reduces the amplitude in the ratio of its transparent portion to its total surface, *i.e.* in the ratio $a/(a+b)$, and hence the intensity of the central image is $\{a/(a+b)\}^2$, if the intensity of the incident light is unity. This determines the value of A.

We now obtain for the intensities of the other images,

$$\frac{1}{n^2\pi^2}\sin^2\left(\frac{n\pi a}{a+b}\right).$$

If $a = b$, the sine factor is zero for all even values of n, so that the spectra of even order disappear, and the intensities of the spectra of odd orders are, in terms of the incident light, $\frac{1}{\pi^2}; \frac{1}{3^2\pi^2} \cdots\cdots \frac{1}{n^2\pi^2}$.

The fraction $1/\pi^2$ represents the maximum intensity which the spectrum of the first order can possibly have in this class of gratings, and shows what a considerable amount of light is lost when a grating is used as an analyser of light. If we desire to make the second order spectrum as intense as possible, we must make a/b equal to $1/3$ or 3, but even in this case, we should only secure little more than two per cent. of the light.

It is instructive to note that the grating reduces the *intensity* of the total light transmitted in the ratio $a/(a+b)$, which is also the ratio in which the *amplitude* of the central image is reduced. The difference between $a/(a+b)$ and $\{a/(a+b)\}^2$ gives the amount of light which goes to form the spectra of higher orders.

67. Gratings with predominant spectra. Rulings of gratings may be devised which concentrate most of the light into one spectrum. Fig. 81 represents the section of such a grating ruled on a reflecting surface. If the oblique portions of the grating lie so that light incident in the direction of the arrow would, by the laws of geometrical optics, be reflected in the direction A_1C_1 then all the rays from each of the oblique portions would be in equal phase at a distant screen HK, placed at right angles to A_1C_1. If, further, the difference

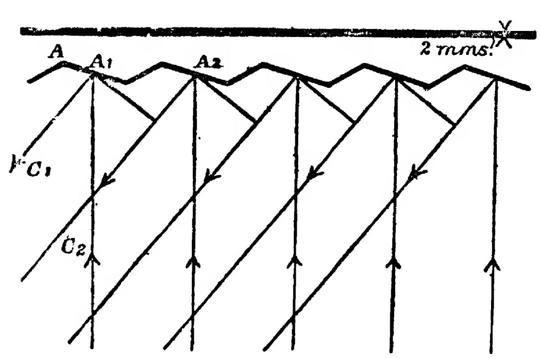


Fig. 81.

in optical length at HK between A_2C_2 and A_1C_1 be a wave-length, there is coincidence of phase between the rays from successive rulings. Hence the amplitude on the screen or at the focus of a lens collecting the parallel rays is the same as if the whole wave-front were reflected in the ordinary way. The resultant amplitude is therefore less than the resultant amplitude of the incident wave, only on account of the contraction in the width of the beam due to obliquity. If θ be the angle between A_1C_1 and the incident beam, it would follow that the intensity of the first order spectrum is $\cos^2\theta$ if that of the incident light be unity. This loss of light is accounted for by the light reflected from the other set of inclined faces. If the ruling is such that the first order spectrum is at an angle of 30° from the normal, three-quarters of the whole light would go to form that spectrum. For normal incidence we have as before, $\sin\theta = \lambda/e$, and the reflecting facets must be inclined at an angle $\theta/2$. The condition for maximum light can only be fulfilled for one wave-length at a time, but a slight tilting of the grating supplies the means of adjustment for any desired wave-length. Transmission gratings may be ruled on the same principle, the condition being that the angles of the inclined facets are such that the incident rays in each little prism formed are refracted along paths at right angles to HK, and that there is a retardation of a wave-length between two corresponding rays A_2C_2 and A_1C_1. Mr T. Thorp has been able to demonstrate the practical possibility of manufacturing gratings of the kind considered. Triangular grooves were cut in a metallic surface, and a layer of liquefied celluloid was allowed to float and solidify over this grooved surface. On removal, the celluloid film showed in transmitted light spectra which were all very weak except that of the first order on one side. HK (Fig. 82) gives the direction of the wave-front of the diffracted wave which carries the maximum intensity for the wave-length λ.

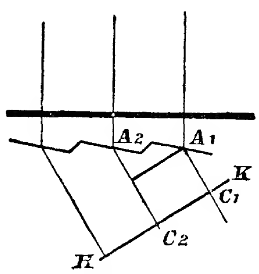


Fig. 82.

68. Echelon gratings. If a reflecting grating were constructed on a principle similar to that of the last article, but subject to the additional condition that rays which go to form a particular spectrum return along the path of the incident light, the spectrum formed by reflexion would contain the whole intensity of the incident light. This consideration leads to Michelson's echelon grating. In Fig. 83 let a number of plates, T_1, T_2, T_3, etc. be placed so that the different portions of a wave-front WF are reflected back parallel

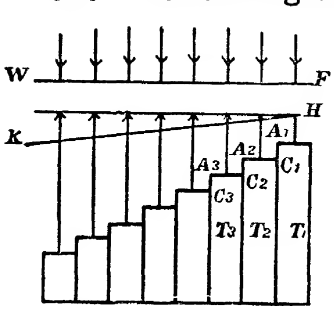


Fig. 83.

to themselves from each of the plates, then if the depths of the steps A_1C_1, A_2C_2, A_3C_3, are all equal to $n\lambda$, a multiple of a wave-length, the reflected beam has intensity equal to the incident beam, neglecting the loss of light at reflexion. For that particular wave-length, there cannot therefore be light in any other direction. The reasoning holds for all those wave-lengths for which the step is an exact multiple of a wave-length, and we may, if n is great, have a great number of maxima of light all overlapping in the same direction.

At a surface HK inclined to WF at a small angle θ, the retardation of successive corresponding rays is $e\theta$, where e is the width of each step. Hence there is coincidence of phase for a wave-length λ' at corresponding points of HK if

$$n(\lambda - \lambda') = e\theta.$$

For the dispersion $\theta/(\lambda - \lambda')$ we thus obtain n/e. But only a very small part of each spectrum is visible because the intensity of light falls off very rapidly to both sides of the normal direction.

At a wave-front parallel to WF, the relative retardation of two waves λ and λ', for the light reflected by the last element, is $Nn(\lambda - \lambda')$ if there is coincidence of phase for light reflected at the first element. Hence equation (2) holds, and the resolving power is Nn, as with ordinary gratings.

A reflecting grating of the kind described would be difficult to construct, but excellent results have been obtained by Michelson with a transmission grating based on the same principle.

A number of equal plates of thickness t are arranged as in Fig. 84. Each part of the beam is retarded by $(\mu-1)t$ more or less than its neighbour. For normal transmission, there is equality of phase everywhere on WF if

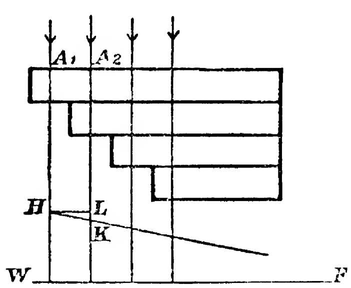


Fig. 84.

$$(\mu - 1)t = n\lambda.$$

If for the wave-length λ', having refractive index μ', the diffracted wave-front is parallel to HK, the phase at H and K must be the same, or

$$(\mu' - 1)t + LK = n\lambda',$$

and if e be the distance between corresponding points A_1, A_2, the angle θ through which the front is turned is LK/e or:

$$\theta = \{n\lambda' - (\mu' - 1)t\}/e$$
$$= \{n(\lambda' - \lambda) - (\mu' - \mu)t\}/e.$$

The angular dispersion is therefore

$$\theta/(\lambda' - \lambda) = \left\{n - \left(\frac{\mu' - \mu}{\lambda' - \lambda}\right)t\right\}\bigg/e. \qquad\qquad (4).$$

If N be the total number of plates, the first minimum of the diffractive image of λ' coincides with the maximum of λ, if the total retardation $Ne\theta$ is equal to λ. Hence multiplying both sides of (4) by Ne, we find:

$$\frac{\lambda}{\lambda'-\lambda} = N\left(n - \frac{d\mu}{d\lambda}t\right) \quad\text{......................(5)},$$

where $\frac{d\mu}{d\lambda}$ has been substituted for $(\mu'-\mu)/(\lambda'-\lambda)$, as only very small variations of μ and λ come into play.

Substituting $n\lambda = (\mu - 1)\,t$, the ratio of the second term on the right-hand side of (5) to the first is $\frac{\lambda d\mu}{d\lambda}\Big/(\mu-1)$, and this for flint glass, and in the centre of the visible spectrum varies between about $-{\cdot}05$ and $-{\cdot}1$. We may therefore say that the value of $\lambda/(\lambda'-\lambda)$ for this form of grating is from 5 to 10 per cent. greater than Nn, but approximately the resolving power is the same as for the nth order of an ordinary grating having a total number N of grooves. Full intensity is only obtained for those wave-lengths for which $t = n\lambda$. But a slight tilting of the grating increases the effective thickness t, and brings any desired wave-length into the best position. The total light is, however, in any case, confined to the immediate neighbourhood of the direction of the incident light, because the width of each element is large compared with a wave-length. It is worth while to discuss this a little more closely. The angular distance between the principal maximum and the first minimum with an aperture e is according to Art. 63, λ/e. We may therefore, disregarding the light which is beyond the first minimum, say that the spectra have appreciable brightness only to a distance λ/e on the two sides of the normal. Consider now that the maximum of the nth order of λ' coincides with the maximum of the $(n+m)$th order of λ when $n\lambda' = (n+m)\,\lambda$. If in (4) we neglect the second term on the right-hand side and for θ substitute $2\lambda/e$ which measures the total angular space within which the light has an appreciable intensity we find $2\lambda = (\lambda'-\lambda)\,n$ or $\lambda' = (n+2)\,\lambda$, which by comparison shows that $m = 2$. No order except n, $n+1$ and $n+2$ can therefore be visible. In the case considered the orders n and $n+2$ would just coincide in position with the places of zero illumination and the central image would contain all the light. As a rule there will be two spectra. As regards intensity of light, the echelon form gets rid of one of the chief difficulties in the use of gratings, as the light must be concentrated almost entirely into two spectra, and we may adjust the grating so that the intensity is practically confined to one spectrum only.

The overlapping of spectra of different orders is, however, a serious inconvenience, for it must be remembered that although for each wave-length there are only two orders visible, the number of the order is different for the different wave-lengths, and the total number of overlapping orders is very great. As an example, consider normal incidence on a grating having its plates of thickness ·5 cm. For a wave-length $\lambda = 5 \times 10^{-5}$, the thickness is 10,000 times the wave-length, so that we should observe a spectrum of the 10,000th order. Coincident with it, and for a slightly differing wave-length, we should have the spectra of orders which are near that number. Thus $n\lambda = \cdot 5$ is satisfied for $n = 8{,}000$, if $\lambda = 6\cdot 25 \times 10^{-5}$. There are therefore 2000 coincident maxima within the range of wave-lengths 5×10^{-5} and $6\cdot 25 \times 10^{-5}$, the former lying in the green and the latter in the orange.

These overlapping spectra must be separated or got rid of. This is done by means of an ordinary spectroscope, which can be used in two ways. In the form of the apparatus as it is most commonly constructed, the light is sent through a train of prisms before it falls on the slit of the echelon collimator. The resolving power of the prisms should be sufficient to exclude all light belonging to the maxima which it is desired to exclude. We may also use a train of prisms to separate the maxima *after* they have passed through the echelon, and this arrangement, which would seem to possess some advantages, was apparently used by Michelson in his first experiments.

69. Concave gratings. That certain gratings possessed a focussing power had been noticed by a number of observers, and the explanation of the fact presents no difficulties, but what previously had always been considered a defect to be avoided, became in the mind of Rowland an object to be desired, and by very perfect mechanical contrivances was made use of to advance spectroscopic research.

It is always possible to construct theoretically the ruling of gratings on surfaces of any shape, such that an image of a spectrum at any desired point shall be formed.

Let A (Fig. 85) represent a point source of light, and let it be desired to form an image of the spectrum of the first order so that all the light of wave-length λ shall be concentrated at B. With A and B as foci, draw ellipsoids such that if P, P', P'' be points on successive ellipsoids,

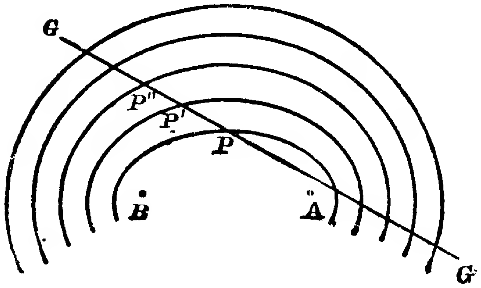


Fig. 85.

$$AP + PB = m\lambda,$$
$$AP' + P'B = (m + \tfrac{1}{2})\lambda;$$
$$AP'' + P''B = (m + 1)\lambda, \text{ etc.}$$

Let GG' be the trace of the surface which it is desired to convert into a grating. The grating intersects these ellipsoids in curves which divide it into Fresnel zones. The light which might reach B from successive zones is in opposition and no luminous disturbance can therefore exist at that point. But if some change be made in the zones, so that the amount of light scattered by alternate zones is either obliterated or at any rate weakened, the plate will act like a zone plate and light will be focussed at B. Ruling lines with diamond point parallel to the lines of division between the zones and at distances equal to the distance between alternate zones, is sufficient to produce the desired effect. As the construction of the zones depends on the wave-length, the spectrum formed has a focus at B for a particular wave-length only. But the adjoining wave-lengths are concentrated into other foci in the neighbourhood. If we desire to produce spectra of higher orders, we may draw the zones so that the sum of the distances of any point from A and B is $m\lambda$, $(m+n)\lambda$, $(m+2n)\lambda$, etc. If a portion of the space filled by each zone so formed is cut by a diamond, so that the corresponding portions of all zones are modified in like manner, a source of light at A produces a spectrum of the nth order at B.

In practice, we are confined to rulings in straight lines on plane or spherical surfaces. We are also unable to rule the lines *accurately* except by means of a screw turned step by step through equal angles. It is Rowland's discovery that gratings with very small aberrations can be made by ruling lines on a spherical surface by means of a screw. In Fig. 86 let A represent a source of light, and B the point at which it is desired to form a spectrum of the nth order. We confine the investigation to rays lying in the plane containing AB and the normal OC of a curved grating GG', C being the centre of curvature. Take OC as axis of X, and the tangent to the grating at O as axis of Y.

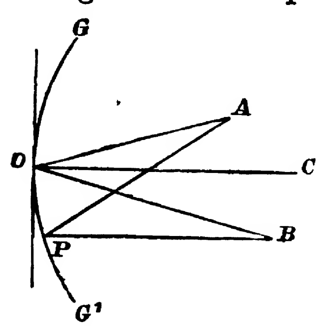


Fig. 86.

Put $OA = r$, $BO = r_1$, $AP = u$, $BP = v$.

If P lies on the edge of the mth zone, and if the nth order spectrum is in focus at B,

$$u + v = r + r_1 \pm mn\lambda.$$

If the distance between successive rulings is such that its projection on OY is constant and equal to e, $y = me$, hence eliminating m,

$$u + v = (r + r_1) \pm \frac{n\lambda y}{e} \quad \text{......................(6).}$$

If this condition could be fulfilled absolutely we should have a perfect image at B. It must be our object now to see how nearly we may satisfy equation (6) in practice.

Writing a, b, for the coordinates of A, a_1, b_1, for those of B, we have

$$u^2 = (y-b)^2 + (x-a)^2$$
$$= r^2 + x^2 + y^2 - 2by - 2ax \quad \text{..............(7)}.$$

If ρ is the radius of curvature of the grating, the equation to the circle in which the grating cuts the plane of the paper gives

$$2\rho x = x^2 + y^2.$$

But our investigation may be made to include gratings, deviating from the spherical shape, so long as the osculatory circle at O has a radius of curvature ρ. We therefore more generally put the equation of the trace of the grating

$$2\rho x = \beta x^2 + y^2 \quad \text{..........................(8)},$$

where β is a numerical constant which is *one* in the case of a sphere. Combining (7) and (8) we obtain by simple transformations

$$u^2 = \left(r - \frac{by}{r}\right)^2 + \left(\frac{a}{r^2} - \frac{1}{\rho}\right) ay^2 + \left(1 - \frac{a\beta}{\rho}\right) x^2 \quad \text{.........(9)}.$$

The second term is of the second order of magnitude as regards y, and the third term of the *fourth* order of magnitude as regards the same quantity. Retaining only quantities of the second order,

$$u = \left(r - \frac{by}{r}\right) + \frac{1}{2r}\left(\frac{a}{r^2} - \frac{1}{\rho}\right) ay^2.$$

Similarly $$v = \left(r_1 - \frac{b_1 y}{r_1}\right) + \frac{1}{2r_1}\left(\frac{a_1}{r_1^2} - \frac{1}{\rho}\right) a_1 y^2.$$

In order that the grating should fulfil its object, it is necessary that at least to this order of magnitude, (6) should be fulfilled. Hence substituting u and v into that equation and putting the factors of y and y^2 equal to zero we obtain:

$$\frac{b}{r} + \frac{b_1}{r_1} = \mp \frac{n\lambda}{e} \quad \text{.........................(10)},$$

and $$\frac{a}{r}\left(\frac{a}{r^2} - \frac{1}{\rho}\right) + \frac{a_1}{r_1}\left(\frac{a_1}{r_1^2} - \frac{1}{\rho}\right) = 0 \quad \text{..............(11)}.$$

The first condition defines the direction in which the diffracted image lies, for if ϕ and θ are the angles which AO and BO make respectively with the normal, $r \sin\phi = b$, and $r \sin\theta = b_1$, and (10) is therefore identical with

$$e(\sin\theta + \sin\phi) = \pm n\lambda.$$

This equation is therefore common to the curved and plane grating. The second condition now gives the *distance* of the diffracted image, for as $r\cos\phi = a$, $r\cos\theta = a_1$, (11) is identical with

$$\frac{\cos^2\phi}{r} + \frac{\cos^2\theta}{r_1} = (\cos\phi + \cos\theta)\frac{1}{\rho}.$$

If θ and ϕ be equal and small, this is the well-known relation between object and image of a concave mirror.

We must now try to see to what order of magnitude we can get rid of aberrations. Leaving out terms of the fourth order equation (9) may be written

$$u^2 = \left(r - \frac{by}{r}\right)^2 + \left(\frac{a}{r^2} - \frac{1}{\rho}\right) ay^2,$$

and hence

$$u = \left(r - \frac{by}{r}\right) + \frac{1}{2}\frac{\left(\frac{a}{r^2} - \frac{1}{\rho}\right)}{r - \frac{by}{r}} y^2 + \text{terms of higher orders.}$$

The term containing y^3 disappears if

$$a\rho = r^2,$$

and as (11) must be satisfied, this involves also

$$a_1\rho = r_1^2.$$

The first equation places the source of light on a circle of radius 2ρ with its centre in the line OC, and the second equation shows that the same circle contains the image B.

Limiting ourselves to this circle for the position of the source, (9) becomes

$$u^2 = \left(r - \frac{by}{r}\right)^2 + \left(1 - \frac{\beta a}{\rho}\right) x^2,$$

and to quantities of the fourth order,

$$u = r - \frac{by}{r} + \frac{x^2}{2r}\left(1 - \frac{a\beta}{\rho}\right),$$

$$v = r_1 - \frac{b_1 y}{r_1} + \frac{x^2}{2r_1}\left(1 - \frac{a_1\beta}{\rho}\right).$$

Comparison with (6) shows that the terms of the fourth order depend on the factor

$$\frac{1}{r}\left(1 - \frac{a\beta}{\rho}\right) + \frac{1}{r_1}\left(1 - \frac{a_1\beta}{\rho}\right) \quad \text{.................}(12).$$

In the position in which Rowland's gratings are generally used $a_1 = \rho = r_1$ and $a = \rho\cos^2\phi = r\cos\phi$. Hence (12) reduces to

$$(1 + \cos\phi)(\sec\phi - \beta)/\rho \quad \text{....................}(13).$$

The terms of the fourth order cannot be got rid of therefore except for a particular value of ϕ. For spherical gratings $\beta = 1$ and the second factor of (13) is small for small values of ϕ, so that the aberration is least important for the spectra of lower orders. It could be corrected entirely for a particular value of ϕ by making $\beta = \sec\phi$, but this would involve departure from a spherical surface.

The outstanding error of optical length for spherical gratings is obtained by restoring in (12) the dropped factor $x^2/2$ or $y^4/8\rho^2$. The error then reduces with $\beta=1$ to

$$\frac{x^2}{2\rho}\sin\phi\tan\phi \quad\text{or}\quad \frac{y^4}{8\rho^3}\sin\phi\tan\phi.$$

The maximum error of optical length may be as much as a quarter of a wave-length without seriously damaging the definition. Hence if y is half the width of the spectrum, we have for the condition of practically perfect definition,

$$\frac{y^4}{2\rho^3}\sin\phi\tan\phi<\lambda,$$

and if $\sin\theta=0$ and $\sin\phi=\frac{n\lambda}{e}$, the greatest value of y which is half the width of the grating should not exceed

$$\left(\frac{2\rho^3 e\cot\phi}{n}\right)^{\frac{1}{4}}=\rho\left(\frac{2e\cot\phi}{n\rho}\right)^{\frac{1}{4}}.$$

The dispersion (Art. 61) of the grating is $n/e\cos\theta$, and for $\theta=0$ is therefore independent of the wave-length. The grating used in such a way that the spectrum appears on its axis forms, therefore, a normal spectrum.

Rowland's method of mounting the grating, which combines the advantages of maximum definition and the formation of a normal spectrum, is shown diagrammatically in Fig. 87. G is the grating and is held by a rigid beam GC of length equal to the radius of curvature of the grating, which carries at its other end the photographic camera HK. AS and BS are two strong beams placed at right angles to each other and carrying rails which support two carriages which can roll along the beams and support in their turn the beam GC which is pivoted on them. The slit is placed at S. As G, S, C lie on a circle of diameter ρ, a luminous source at S will always have its image at C, when the proper position of the beam GC has been found.

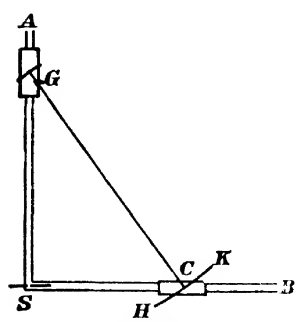


Fig. 87.

As the beam is rolled along the rails, successive wave-lengths and successive spectra make their appearance in proper focus at C. For a given position of the beam, the focus for the different wave-lengths lies on a circle of diameter ρ, and the photographic plate HK must therefore be bent to a curvature equal to twice that of the grating. The angle CGS is the angle called ϕ above, and as $\theta=0$,

$$e\sin\phi=n\lambda,$$

and as

$$SC=\rho\sin\phi,$$

it follows that

$$SC=\frac{n\lambda\rho}{e}.$$

The beam SB may therefore be divided into a scale of wave-lengths by equal divisions, and the wave-length which occupies the centre of the field at C may be read off directly on that scale.

A complete discussion of the theory of the concave grating which we have in great part followed here, has been given by Runge, and published by Kayser (*Spectroscopie*, Vol. I. p. 400). The same volume contains valuable information on the methods of adjustment and on the literature of the subject. It should however be mentioned that though later investigations have simplified the analysis, the essential points of the theory are all contained in Rowland's* original papers.

70. Measurement of wave-length. Plane gratings allow us to measure the length of a wave of light (λ). In Fig. 88, C represents a

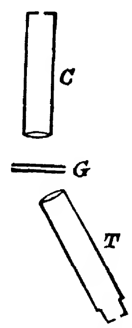


Fig. 88.

collimator which admits the light through a narrow slit. The source must be one sending out nearly homogeneous radiations, as *e.g.* a tube filled with a vapour under reduced pressure and rendered incandescent by the electric discharge. The light is allowed to fall on a grating at G and is observed by means of a telescope T. If the axis of the telescope coincides with the direction of a maximum of light in the diffracted beam, we have the relation

$$e(\sin\phi + \sin\theta) = n\lambda,$$

where the letters have the same meaning as in Art. 60. If the incident wave-front coincide with the plane of the grating, $\phi = 0$ and θ becomes the deviation. If the grating be used in minimum deviation $\phi = \theta$ and the deviation is $2\theta = \frac{n\lambda}{e}$. The deviation being capable of very accurate determination, the wave-length is found

* "On Concave Gratings for Optical Purposes"; *Phil. Mag.* XVI. p. 197 (1883). See also J. S. Ames' "The Concave Grating in Theory and Practice"; *Phil. Mag.* XXVII. p. 369 (1889).

directly, when the distance between the lines of the grating is known.

When high accuracy is required the grating ceases to be an efficient means of determining wave-lengths, on account of the difficulty of avoiding irregularities in the ruling. If these, as is often the case, are of a periodic character, the average value of e ceases to determine completely the position of the maximum of light.

An important aspect of the problem comes into view when it is inverted. In measuring a wave-length we are comparing the length of a wave of light with our standards of linear measurement. These standards are arbitrary; their permanence depends on the preservation of a metallic rod, which is liable to vary in length owing to temporary or permanent causes. It is therefore not surprising that a proposal —first made by Lamont in 1823, and subsequently emphasized by Maxwell—to express the common standards of length in terms of that of a homogeneous radiation, should have received serious attention when experimental methods were sufficiently advanced. This state was reached when Michelson had introduced his interference method, which for this purpose is much more accurate and reliable than the grating.

A comparison of the standard metre, preserved at the Bureau International des Poids et Mesures at Meudon, near Paris, with the radiations of three Cadmium lines, was successfully accomplished by Michelson in 1894. A certain number of intermediate optical standards were constructed, each containing as its essential part two

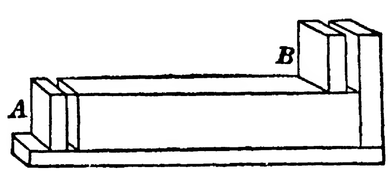


Fig. 89.

mirrors A and B (Fig. 89), with the necessary adjustments to allow their silvered surfaces to be placed parallel to each other. The distance between the planes of the silvered surfaces of A and B

determines their optical length. Nine standards were used, having distances between the mirrors respectively of 10 cms., $\frac{1}{2} \times 10$ cms., $\frac{1}{4} \times 10$ cms., and so on, the smallest being equal to $\frac{1}{256} \times 10$ cms. or a little less than half a millimetre. Their relative lengths were determined by comparing each of them with the one having nearly double or half the length. The deviations from the exact commensurate ratio is determined by the interferometer. The disposition of the mirrors of the instrument has to be adapted for the purpose, an additional reflecting surface *H* being introduced, and the front surface of *A* (Fig. 90) taking the place of *c* in Fig. 44.

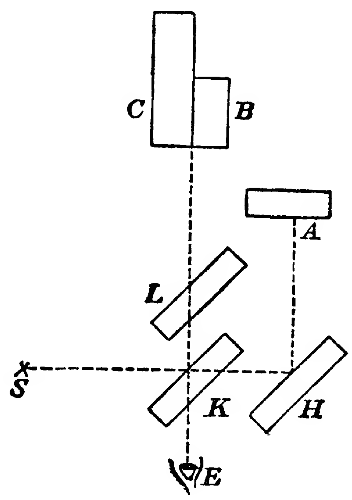


Fig. 90.

The eye observing at *E* through a telescope sees the image of *A*, reflected at *K* superposed on the images, seen directly, of the four mirrors on *B* and *C* which lie side by side and above each other. Under suitable conditions of distance, interference fringes are then seen. The mirrors *A* and one of the standards *C* are placed on carriages which by means of two parallel screws can be moved independently backwards and forwards in such a manner that their surfaces remain absolutely parallel.

Let it be required to measure in wave-lengths of a particular radiation the difference between the length of the standard *C* and the double of the length of the standard *B*. The following operations are necessary.

1. Place the surface of reference A at the same optical distance as the front surface of C, determined by the plane of its lower mirror: turn that mirror slightly about a vertical axis. With white light, vertical interference fringes will appear projected on the lower surface of C, and the dark central fringe which marks the position of zero difference of path may be placed at the edge of C which is nearest to B. Now displace the mirror B parallel to itself until vertical fringes also appear on its lower mirror. Adjust the distance and tilt the mirror until the fringes are spaced equally with those on B and the central dark fringe lies on that edge of B which is nearest to C. At the end of this operation the lower part of the field contains vertical fringes both on the left and right side, and the two central black fringes are almost in contact with each other.

2. Move the plane of reference A parallel to itself until its relative position to the further and upper mirror of B is the same which it previously had to its nearer and lower mirror. The dark fringes will now appear only on the right and upper quadrant of the field.

3. Move the standard B back through its own length so that its front mirror now occupies the same position relative to A which it had at the end of the first operation. The dark fringes are now seen on the right and lower quadrant of the field.

4. Move the plane of reference back until it has the same relative position to the upper mirror of B which it had at the end of the second operation. The dark fringes now appear again in the right and upper quadrant of the field. But if the standard C has exactly double the length of B the fringes will also be seen in the left upper quadrant. If the standards have been sufficiently well adjusted to begin with, this will be the case, but the central fringe will be slightly displaced. This displacement in terms of the number of fringes (say of the sodium radiation) gives the required difference in wave-lengths. Fractions of wave-length may be measured by slightly tilting the compensator L,

which moves the fringes so that they can be made to occupy any desired position.

The length of the shortest standard must be determined in terms of the wave-length of some homogeneous radiation, the red line of Cadmium being suitable for the purpose. To perform this part of the work, let B be the standard to be measured and place some other standard, say C, at such a distance that with proper adjustment as regards parallelism, rings of suitable diameters appear on one of its mirrors. At the same time the front surface of B is brought to the same optical distance as A and slightly inclined so that the vertical fringes appear as before. The mirror A is now moved back so slowly that the circular fringes on B may be counted as they contract and disappear at the centre. The vertical fringes on the lower mirror of B disappear as the optical distance of A is increased, but they reappear on the upper mirror in exactly the same position as soon as A has moved through the length of the standard. The whole number of rings that have disappeared together with a fraction which can be measured by the compensator gives the required number of wave-lengths. It is of course necessary to use homogeneous light for this ring system, while at the same time white light is necessary to observe the straight line fringes under the best conditions.

The next operation consists in obtaining a length equal to ten times that of the longest standard of 10 cms. This is accomplished by moving the standard back ten times through its own length. The previous description suffices to show that this can be done by keeping A stationary while moving the standard back and observing the fringes to appear first on its lower and next on its upper mirror. The mirror A is next moved back until the fringes appear again on the lower mirror of the standard, and this is repeated ten times. Finally, the whole distance through which the standard is moved is compared with the metre. For this purpose the standard carries a side-piece having a line marked on it which can be brought into the same field of view of a microscope as a similar mark attached to the ends of a metre measure. The above gives, in outline, the method, which was admirably carried out by Professor Michelson. For all details the original paper* should be consulted. As the result of this work the wave-lengths of three of the Cadmium lines referred to air at a temperature of 15° C. and a pressure of 760 mm. were found to be:

red line: $\lambda = 6438{\cdot}4722$ Å,
green line: $= 5085{\cdot}8240$,
blue line: $= 4799{\cdot}9107$.

The symbol Å denotes the unit now generally adopted and called an Ångström, which is equal to 10^{-8} cms. According to a suggestion of Johnstone Stoney it is also sometimes called a tenth-metre.

An independent comparison of the metre with the wave-length has recently been made by Messrs Benoit, Fabry and Perot†. Important advantages were gained in this new determination owing to the greater simplicity of the apparatus and to the shorter time occupied in the measurement: this is important because the chief danger to the accuracy of the work lies in the changes of temperature and the consequent changes in the length of the standards during the measurement.

The interference plates of Fabry and Perot have already been described (Art. 43). For the purpose of this research five standards

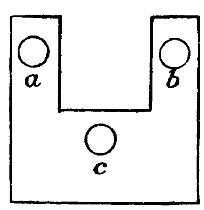


Fig. 91.

were constructed, each twice as long as the next shorter one. The vertical section of the frame of the standards made of invar metal had the form shown in Fig. 91. Three invar screws *a*, *b*, *c* fixed to

* Bureau International des Poids et Mesures: *Travaux et Mémoires.*

† *Transactions of the International Union of Solar Research.*

both extremities of the frame are provided with hemispherical polished heads. Against these the silvered surfaces of glass plates which have a surface of 5 square cms. are pressed by three springs. The surfaces of the heads may be ground until the plates placed against them are almost exactly parallel and the final adjustment may be then made by altering the tension of the springs which press the plates against the supports. The distance between the silvered glass plates in the successive standards was: 6·25, 12·5, 25, 50 and 100 cms. As in the case of Michelson's determination, each had to be compared with the next, and the difference between its length and the double of the length of the previous one had to be measured. But the process here is much simpler as the position of the standards remains fixed the whole time. The criterion that one of the air-plates is exactly double the length of the other is found by means of the bands which appear with white light when the relation is approximately established. The bands are of the same nature and follow the same law as Brewster's bands.

As it is impossible to obtain accurately the theoretical relationship of 2 : 1 in the standards, means must be found to determine the error. For this purpose the light after passing through the standards is sent through a thin wedge of air, included between two silvered glass plates, which serves as compensator. If of the two plates under comparison one has nearly twice the thickness of the other, a ray which has passed through the thicker one and been reflected twice backwards and forwards in the thinner one has traversed an optical path very nearly equal to that of another ray which has been reflected once backwards and forwards inside the thicker one. If E and E' be the distances between the silvered plates of the standards respectively the difference in path to be corrected is $\pm(4E' - 2E)$. If $2E'$ is greater than E, the central fringe appears on the wedge at a point where its thickness is e, if the ray which has traversed $2E$ has received an additional retardation $3e$ (by means of two internal reflexions) while the ray which has traversed $4E'$ passing straight through the wedge has an additional path of e only. We have then, where the central fringe appears in white light,

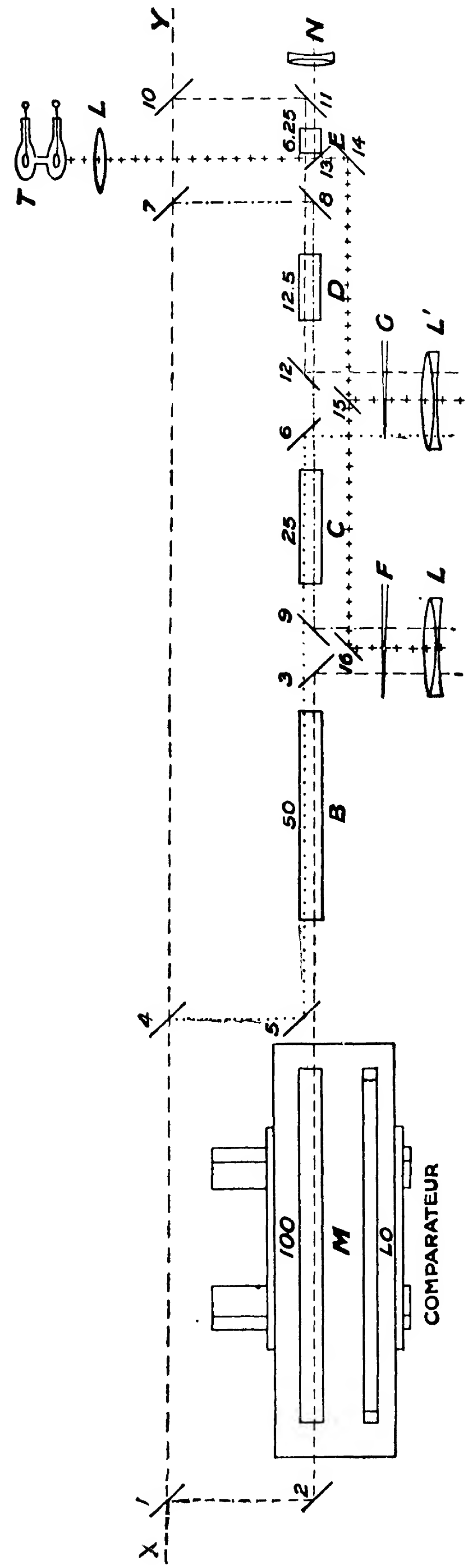
$$4E' + e = 2E + 3e,$$

$$\therefore \ e = 2E' - E.$$

It is clear that the fringes must be observed by an eye or microscope which is focussed on the wedge. In white light the position of the central fringe can be accurately observed if the standards have previously been placed exactly parallel to each other. The manner in which the wedge has to be standardized so that the quantity e may be determined in wave-lengths presents no difficulties.

The complete arrangement of the standards placed ready for comparison is shown diagrammatically in Fig. 92. All standards are placed in a line and carefully adjusted so as to be accurately parallel. The comparison can then be made quickly without moving any of them. If *e.g.* the standard M, having a length of one metre, is to be compared with B, having a length of 50 cms., white light coming from X is reflected by the mirrors 1 and 2, and traverses M and B, the mirror 5 being moved out of the way. It is then reflected by the mirror 3 and passes through the compensating wedge F, being finally observed through the magnifying glass L. Inspection of the figure shows how the different mirrors are used when comparing successively the different standards. The two remaining operations consist in measuring the shortest standard in terms of the wave-length and in comparing the longest with the metre measure, marked LO in the figure. As regards the length of the 6·25 cm. standard, this can be determined once for all to the nearest whole wave-length, as changes of temperature only affect the fraction, which is found in each comparison by measuring the diameter of the first ring.

For this purpose the tube T is used which produces cadmium light of sufficient purity, and the circular fringes are observed at each determination through the telescope N The mirrors 14 and 15 or 16 serve to observe the fringes of cadmium light in the wedges F or G.



Y
N
10
11
6.25
E
13
14
T
L
7
8
12.5
D
G
L'
12
15
6
25
C
F
L
9
3
16
50
B
4
5
100
M
LO
COMPARATEUR
1
2
X

Fig. 92.

Though these have been carefully calibrated beforehand, the calibration is only used to determine the complete number of wave-lengths in the retardation e, the fraction being determined at the time of observation by noting the position of the cadmium fringes.

The length of the metre LO made of invar was accurately known by means of direct comparisons made with the standard metre. It was placed during each optical comparison side by side with the standard M in a "comparator" which allowed a couple of reading microscopes, kept at a fixed distance from each other, to be brought quickly either above the marks which define the metre or above two marks made in the upper horizontal surface of the glass plates of the standard M. The only remaining measurement consists in determining the small horizontal distance of these marks from the silvered surface of M which served during the comparison with the 50 cms. standard marked B in the figure. For the description of the method employed the reader is referred to the original paper.

As a result of the measurements the wave-length of the red cadmium line was found to be

$$\lambda = 6{\cdot}4384696 \times 10^{-5} \text{ cms.}$$
$$= 6438{\cdot}4696 \text{ Å}.$$

The difference of about one part in 2·5 millions between this number and that found by Michelson is mainly accounted for by the moisture contained in the air, Michelson's number referring to air containing an undetermined amount of water vapour. If the correction to dry air be estimated and a small alteration made to render the scales of temperature used identical with each other, Michelson's value becomes 0·643847000 which is practically identical with the number obtained by Messrs Benoît, Fabry and Perot. This number may now be taken to fix once for all the scale of wave-lengths to be used in spectroscopic measurements. It defines the Ångström, and whatever slight error there may still remain would only cause the Ångström unit to differ from 10^{-10} cms. by less than one part in a million.

Independently of their application to the establishment of absolute standards, Fabry and Perot's silvered air-plates are likely to render important services in determining relative values of wave-lengths; a convenient method for the purpose has been described by Lord Rayleigh*.

71. Historical. Joseph Fraunhofer (born March 6, 1787, died June 7, 1826) was engaged from an early age in a glass manufacturing works, and became specially interested in the construction of telescope lenses. He recognized the fact that their improvement, especially as regards achromatism, depended on an exact determination of refractive indices, and that the chief difficulty in that determination lay in the difficulty of obtaining homogeneous radiations which could serve as standards. The sodium flame was made to serve as one kind of radiation, and in using sunlight he discovered that nature had placed standard radiations at his disposal. The spectrum of the sun was seen to be traversed by dark lines—now called Fraunhofer lines—which marked the position of homogeneous radiations by a deficiency in radiance just as well as could have been done by an increase in it. To the earlier observation of these lines by Wollaston no importance had been attached, because it had not been recognized that their position was invariable and independent of the mode of observation. A few years before his early death, Fraunhofer was led to the study of diffraction effects and constructed the first gratings, by stretching fine wires between two screws having narrow threads, and also by ruling lines with a diamond point on a glass surface. He used these gratings for the determination of the wave-length of the principal Fraunhofer lines.

Table VI. gives in Å. U. of 10^{-8} cms. the wave-lengths as obtained by Fraunhofer and subsequent observers.

Not much progress could be made in improving the accuracy of wave-length determination until the manufacture of gratings was improved. Those made by Nobert towards the middle of last century

* *Collected Works*, Vol. v. p. 312.

obtained considerable reputation, and Ångström (born Aug. 13, 1814, died June 21, 1874, in Upsala) constructed an Atlas of the Solar Spectrum with one of Nobert's gratings, which for a considerable time remained the standard to which all wave-lengths were referred.

TABLE VI.

Solar line	Fraunhofer 1823	Ångström 1868	Rowland 1887
C	6561	6562	6563
D	5890	5892	5893
E	5268	5269	5270
F	4859	4861	4861
G	4302	4307	4308
H	3963	3968	3969

Lewis Morris Rutherford, an amateur astronomer, and lawyer by profession, ruled gratings, by means of an automatically acting dividing engine, which were considerably better than any previous ones. He was the first to rule gratings on metal, which being softer than glass did not destroy the ruling edge of the diamond to the same extent. Most of his gratings were made about the year 1880.

Henry A. Rowland (born 1848, died 1901) effected still greater improvements. An essential portion of machines intended to rule gratings is the screw, which should be as free from errors as possible. Slight accidental displacements of the lines, so long as they are not systematic, and especially not recurring periodically, are not of serious importance. Rowland's first achievement consisted in the making of a screw more perfect than any made before. The following passage taken from his article on "Screw" in the *Encyclopaedia Britannica* gives an idea of the method he adopted:

"To produce a screw of a foot or even a yard long with errors not exceeding $\frac{1}{1000}$th of an inch is not difficult. Prof. Wm. A. Rogers, of

Harvard Observatory, has invented a process in which the tool of the lathe while cutting the screw is moved so as to counteract the errors of the lathe screw. The screw is then partly ground to get rid of local errors. But, where the highest accuracy is needed, we must resort in the case of screws, as in all other cases, to grinding. A long solid nut, tightly fitting the screw in one position, cannot be moved freely to another position unless the screw is very accurate. If grinding material is applied and the nut is constantly tightened, it will grind out all errors of run, drunkenness, crookedness, and irregularity of size. The condition is that the nut must be long, rigid and capable of being tightened as the grinding proceeds; also the screw must be ground longer than it will finally be needed so that the imperfect ends may be removed."

CHAPTER VII.

THE THEORY OF OPTICAL INSTRUMENTS.

72. Preliminary discussion. There is a limit to the power of every instrument, due to the finite size of the wave-length of light. According to the laws of geometrical optics, the image of a star formed in a parabolic mirror should be a mathematical point, and if this were the case the sole consideration to be attended to in the construction of optical instruments would be the avoidance of aberrations. According to the wave theory of light, however, the image of a point source is never a point, however perfect the instrument may be in other respects, and the longer the wave-length the more does the light spread out sideways from the geometrical image. It is therefore useless to try to avoid aberrations beyond a certain point, and it becomes a matter of primary importance to define the natural limit of the power of an instrument, so as to be able to form a clear idea as to how far the optician may usefully spend labour in the refinement of his surfaces.

Let a wave divergent from a point source A (Fig. 93) be limited by an aperture SS' in a screen, and let the light transmitted through this aperture be no further obstructed in its passage by any perforated screens, but pass entirely through lenses, or be reflected or refracted in any manner until ultimately the wave surfaces become portions of spheres concave towards a point P.

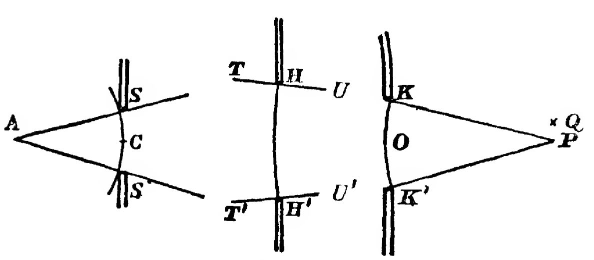


Fig. 93.

It will be necessary to calculate the amplitude in the light in the neighbourhood of P, and a preliminary proposition will help to simplify the problem. Trace the rays AS, AS', limiting the beam, according to the laws of geometrical optics, and let TU, $T'U'$ be portions of these rays. Place screens at KK' or HH' with apertures just sufficient to transmit these rays, or, in other words, let the edge of the aperture KK' or HH' coincide with the geometrical shadow of the opaque portions of the screen SS'. The proposition to be proved is, that the introduction of these screens does not alter the distribution of light in the neighbourhood of P, and that the screen SS' may be replaced by either of them, leaving all the amplitudes near P as they were. The truth of the proposition depends on all portions of the wave surface passing through KK' contributing equally to the amplitude of P, as P being a point of convergence of the rays, its optical distance to any point of KK' is the same. The screen KK' obliterates only the waves which have spread out laterally before they have reached the plane of the screen. The portions so obliterated are a very small fraction of the light forming the image at P which is due to the combined action of the complete wave. The same is true for the resultant amplitude at Q so long as the aperture KK' only contains a small number of Fresnel zones drawn from Q as centre.

In order that students should not be misled into an erroneous application of this proposition, we give an example where it fails. SS_1 (Fig. 94) is a screen limiting a parallel beam of light, RS being the edge of the geometrical shadow. SS_1 cannot here be replaced by a screen TT_1 giving the same geometrical shadow because in this case equal parts of the wave-front do not equally contribute to the amplitude at Q. Tracing Fresnel zones from Q, the loci of the division between two zones are parabolas (Art. 52). The parabola QS traces the limiting zone for the screen SS_1, while for TT_1 the limiting curve would be a different parabola QT. If the angular space TQS includes an odd number of zones, the change of position of the screen from SS_1 to TT_1 would cause a difference in amplitude equal to that of a complete zone, so that a maximum of light might be changed into a minimum or vice versa.

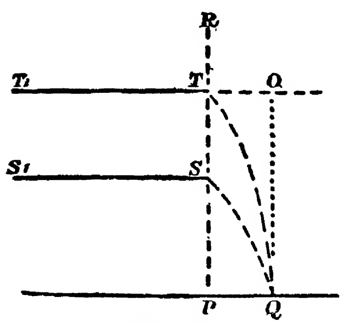


Fig. 94.

73. Image formed by a Lens. It is convenient to imagine the beam to be now limited by a diaphragm just inside the lens which concentrates the light at F. The traces of the wave-fronts are circles with F as centre, and if $D=2R$ is the diameter of the lens, ρ the distance of any point P from F, and $f=GF$,

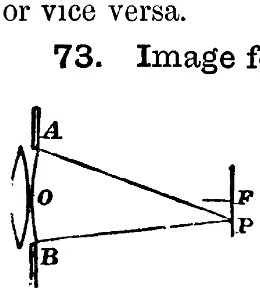


Fig. 95.

$$AP^2=f^2+(R+\rho)^2,$$
$$BP^2=f^2+(R-\rho)^2;$$
$$\therefore\ AP^2-BP^2=4R\rho,$$

and ρ being very small compared with f,

$$AP - BP = \frac{2R\rho}{f} \quad \text{.........................(1).}$$

If we were only to consider rays in the plane of the paper, then light at P would be destroyed by interference if

$$\frac{2R\rho}{f} = n\lambda \quad \text{.............................. (2),}$$

and bands of maximum brightness would appear where $\frac{2R\rho}{f} = (n + \tfrac{1}{2})\lambda$, approximately. If we imagine the figure to revolve round the axis OF, the luminous appearance in the plane through F, at right angles to the axis, would be a luminous disc fading outwards until the intensity becomes zero when $\rho = f\lambda/D$. This disc would be surrounded by dark and bright rings, the brightest parts of the rings corresponding to the distance $\rho = (n + \tfrac{1}{2}) f\lambda/D$.

Owing to the rays which do not lie in the plane of the paper the destruction of light takes place at a distance somewhat greater from F than that given by the above approximate calculation.

Sir George Airy* was the first to solve the problem of the distribution of light in the image of a point source. His solution is expressed in the form of a series, while Lommel gave it subsequently in terms of Bessel functions. The main effect however is correctly represented by the above elementary considerations. The diffraction image is a disc surrounded by bright rings, which are separated by circles at which the intensity vanishes.

If we write $$\rho = m\frac{f\lambda}{D} \quad \text{.............................. (3),}$$

the values of m for the circles of zero intensity are given in the following Table. They differ very nearly by one unit, but instead of being integers, as the approximate theory would indicate, approach a number which exceeds the nearest integer by about one quarter.

TABLE VII. *Dark rings.*

Order of ring	m	Total light outside dark circle
1	1·220	·161
2	2·233	·090
3	3·238	·062
4	4·241	·048
5	5·243	·039
6	6·244	·032

* *Trans. Camb. Phil. Soc.*, v. p. 283 (1834).

The third column of Table VII. gives the amount of light lying outside each ring. Thus the first number ·161 indicates that ·839 of the total light goes to form the central disc and the difference between the first and second number gives the fraction of the total light which forms the first ring. These differences are recorded in the last column of Table VIII. which is mainly intended to give the values of m for the circles of maximum illumination and the corresponding intensities. The third column contains the intensity at the maximum in terms of the central intensity.

TABLE VIII. *Bright Rings.*

Order of disc or ring	m	Maximum intensity in terms of central intensity	Fraction of total light in disc or ring
1	0	1	·839
2	1·638	·01745	·071
3	2·692	·00415	·028
4	3·716	·00165	·015
5	4·724	·00078	·009
6	5·724	·00043	·006

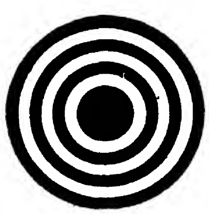


Fig. 96.

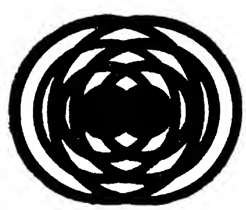


Fig. 97.

Fig. 96 gives in diagrammatic form the relative sizes of the central disc and the first three rings. Fig. 97 shows the images of two sources of light placed at such a distance apart that the centre of the bright disc of one falls on the first dark ring of the other.

74. Resolving Power of Telescopes. It has long been known to all astronomers working with high powers, that the image of a star in a telescope has the appearance roughly represented in Fig. 96, and it is a matter of experience that a close double star may be recognized as such when the relative position of the stars is not closer than that represented by Fig. 97. This allows us to calculate the angular distance between the closest double star which the telescope can recognize as such.

The radius of the first dark ring being ρ and the focal length of the telescope being f, the angle θ subtended at the centre of the object glass by two stars which occupy such a position that the centre of the diffraction image of one falls on the first dark ring of the other is ρ/f, which by (3) gives

$$\theta = 1{\cdot}22\,\lambda/D \quad\quad (4).$$

This is the angular distance between the stars when they are on the point of resolution and no subsequent refraction of light through lenses can alter this angle. The images may be enlarged but the rings and disc are always enlarged in the same ratio. This is an important fact which may be more formally proved in this way: If the rays crossing at any point of the diffraction image Q (Fig 98) are brought by a lens or system of lenses to cross again at a point Q', the optical distance from Q to Q' along all paths must be the same, and hence the retardation of phase between any two rays at Q is accurately reproduced again at Q'. If there is neutralization at Q, there must also be neutralization at Q'. As Q' is the geometrical image of Q, the diffraction pattern in the plane of Q' must be the geometrical image of the diffraction pattern in the plane of Q. Our result may therefore be applied to eye observations through a telescope, the plane of Q' representing the plane of the retina.

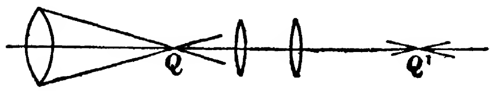


Fig. 98.

It appears from the above that the power of a telescope to resolve double stars is proportional to the diameter of the lens. This result, which agrees with the facts, depends on the wave-theory, for if the rays were propagated by the laws of geometrical optics, the size of the object glass would be immaterial, while the angular separation could be increased at will by a suitable magnifying arrangement. We also see that the smaller the wave-length, the more nearly are the laws of geometrical optics correct.

To resolve stars at an angular distance of 1 second of arc ($4{\cdot}84 \times 10^{-6}$ in angular measure), we should for $\lambda = 5 \times 10^{-5}$ require a linear aperture of

$$D = 1{\cdot}22 \times \frac{5 \times 10^{-5}}{4{\cdot}84 \times 10^{-6}} = 12{\cdot}6 \text{ cms.}$$

Hence the angular distance θ in seconds of arc, which an object glass of diameter D can resolve, is

$$\theta = \frac{12{\cdot}6}{D}.$$

The Yerkes telescope with an aperture of 100 cms. should be able therefore to resolve two stars at a distance of one-eighth of a second of arc. This calculation is based on the supposition that the whole of the light which passes through the telescope enters the eye. By a well-known law which will be proved in Art. 78, the magnifying power of a telescope is equal to the ratio of the widths of the incident and emergent beams. If the width of the emergent beam is greater than the greatest width p which is capable of entering the pupil of the eye, the full aperture is not made use of. Hence to obtain full resolving power the magnifying power of a telescope should be not less than D/p. If it is less, the rays entering the outer portion of the telescope lens do not enter the eye at all and may as well be blocked out altogether, thus reducing the aperture to its useful portion.

75. Resolving Power of the Eye. We may apply equation (4) to the case of two stars or other point sources which are looked at directly by the eye. It might be thought that a complication arises owing to the fact that the wave-length of light in the vitreous humour, which is the last medium through which it passes, is not the same as the wave-length in air, but this makes no difference. To show this let Fig. 99 represent diagrammatically a beam of light entering the media of the eye. If a plane wave-front passes through an aperture AB of such size that the beam passing through it may just enter the pupil of the eye, the first dark ring of the diffraction images passes through Q when the difference in optical lengths from A to Q exceeds by $1{\cdot}22\lambda$ that from B to Q. Also a wave-front parallel to $W'F'$ has the centre of its diffraction image at Q when the optical distance from all points of its plane to Q is the same, hence AT must be equal to $1{\cdot}22\lambda$, and the angle between AB and $W'F'$ is measured by AT/AB or

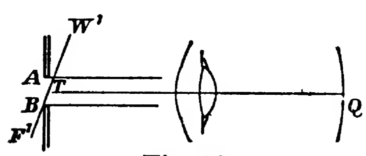


Fig. 99.

$$1{\cdot}22\frac{\lambda}{p}.$$

Here λ is the wave-length measured in *air*.

The width of the pupil is variable, but with light of medium intensity such that p is about 3 mm. (the actual opening of the pupil will be less, owing to the convergence produced by the cornea), two small point sources of light should be resolvable by the eye when at an angular distance of 42″. Helmholtz gives for the experimental value of the smallest angular distance perceptible by the eye the range between 1′ and 2′, which would show that with full aperture of the pupil, our sense of vision is limited rather by the optical defects of the eye and physiological causes than by diffraction effects.

76. Rectangular Apertures. If the surface of the telescope is covered by a diaphragm having a rectangular aperture, the distribution of light is more easily calculated, and may be expressed accurately in a simple form. Let a parallel beam of light of unit intensity fall on a lens having a focal length f, the lens being covered by a diaphragm with a rectangular aperture of width a and length b. Introduce a system of coordinates into the focal plane having its origin at the geometrical image, the direction x and y being parallel to a and b respectively. The complete investigation, which it is not necessary to reproduce here, shows that the illumination I at any point of the screen is given by

$$I = \frac{a^2b^2}{f^2\lambda^2}\frac{\sin^2\alpha}{\alpha^2}\frac{\sin^2\beta}{\beta^2} \quad \text{.......................(5)},$$

where $\alpha = \pi ax/f\lambda$ and $\beta = \pi by/f\lambda$.

The same equation holds good for the image of a point which is either sufficiently far away or is provided with a collimator lens. If the point is moved up or down through a distance e' in a direction parallel to y, its image is moved up or down through a distance e, if e/e' be the magnification of the arrangement. We may now calculate the distribution of light in the image of a slit placed parallel to the axis of y. A point of the slit which is at a distance y' from the centre will produce an illumination at the centre of the image which is equal to that produced by the centre of the slit at a distance $y = y'e/e'$ from the centre of the image. The total illumination at the centre of the image, due to a length $2e'$ of the slit, is therefore

$$\int_{-e}^{+e} Idy.$$

If e be sufficiently large we may substitute $\pm\infty$ for $\pm e$ at the limits of the integration and make use of the known value of the definite integral

$$\int_{-\infty}^{+\infty}\frac{\sin^2 v}{v^2}dv = \pi.$$

We thus find that the illumination at the image is proportional to

$$I' = \frac{a^2b}{f\lambda}\frac{\sin^2\alpha}{\alpha^2},$$

where α has the same value as before. It follows that the total amount of energy which is transmitted in unit time through a small surface s of the image is $\kappa sI'$, where κ is a constant which may be determined as follows. If a ribbon of unit width be cut out transversely to the image, the total amount of energy transmitted through the ribbon is

$$\kappa\int_{-\infty}^{+\infty} I'dx = \kappa ab.$$

If E denote the amount of light from unit length of the source transmitted through unit surface of the rectangular aperture, and m the magnifying power, the total amount of light per unit length of the image is Eab/m. Hence $\kappa = E/m$.

77. Luminous Surfaces. The image of a surface bounded by a straight edge may be calculated from the above. Dividing the surface into narrow strips parallel to one of the edges, each strip will have a diffraction image in which the illumination varies as represented in Fig. 70, and at each point of the image we should have to add up

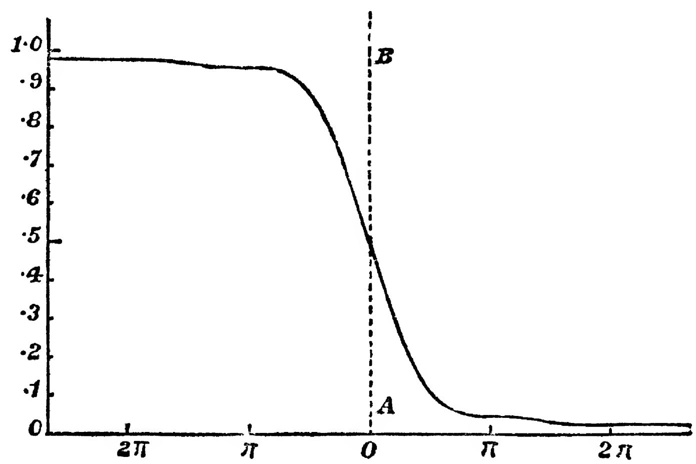


Fig. 100.

the effect due to each strip. It is easy to see that at the geometrical image of the edge, the illumination is half that observed at some distance outside the edge, where the illumination is uniform, for when two similar surfaces are placed against each other with their edges in contact a uniformly illuminated image is obtained, and each half must contribute equally to the illumination at the dividing line. The blocking out of one half must consequently halve the illumination. At other points it can only be expressed in the form of definite integrals or calculated by means of a series. The intensities are plotted in Fig. 100. The dotted line AB marks the edge of the geometrical image of the surface. The intensity at that point is ·5, and falls off rapidly towards the outside of the image.

When a telescope is used to examine an extended surface such as the moon, the problem consists in interpreting differences in an unequally illuminated surface. Details which are as near together as two stars when at the point of optical separation may be indistinguishable on an extended surface. To distinguish clearly between two small features in an extended surface it is generally necessary to increase the resolving power until the central discs of their images stand quite clear of each other.

For a given distance from the geometrical edge the intensity is less than at the same distance from the image of a narrow aperture. Hence, as has been pointed out by Wadsworth, the images of two surfaces may be closer together than the images of the slit without their images becoming confused.

The points marked π and 2π on the horizontal line of Fig. 100 represent the places where the first two minima of light would occur in the image of a narrow slit coincident with AB.

78. Illumination of the image of a luminous surface. The resultant energy which leaves a luminous surface is the same in all directions for equal cross-sections of the beam. As with a given small surface S, Fig. 101, the cross-section of the beam varies as the cosine of the direction angle θ, the intensity of radiation sent out by a surface S is proportional to $\cos\theta$, but for small inclinations to the normal, we may take the radiation to be independent of the direction. If an image of a surface S is to be formed, the illumination of the image must be proportional to the amount of light which the luminous surface sends through the optical system. If all the light which passes through the first lens passes also through the other lenses, this is proportional to the surface S, and to the solid angle ω subtended by the lens at a point of S. We may therefore write for the light passing through the optical system $IS\omega$, where I solely depends on the luminosity of the surface. If s is the size of the image of S, and if the image is such that the illumination is uniform, the brightness of the image is equal to $IS\omega/s$.

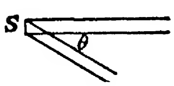


Fig. 101.

We shall first consider the case that the linear dimensions of s are such that the diameter of the diffraction disc may be neglected in comparison with it, so that we may find the relation between S and s by the laws of geometrical optics.

Let LL' and MM' be the wave-fronts diverging from P and converging to the image Q respectively, and imagine a second wave-front RR' slightly inclined to the first, to diverge from P'. If $PO = P'O$, the second wave-front may be obtained by turning LL' about O through a small angle θ. The optical length from P' to L has been increased in the change, by the quantity RL, and the optical length from P' to L' has diminished by the same amount. The optical lengths from L to M and $L'M'$ have not been altered (Art. 23). Hence if Q' is the image of P' the optical length $M'Q'$ must differ from MQ' by $2RL$, the total length from P' to Q' being the same whether measured by way of LM or by way of $L'M'$. It follows that to obtain Q' we must turn round the wave-front $M'M$ through such an angle that $HM = 2RL$. If D is the width of the beam at LL' and d the

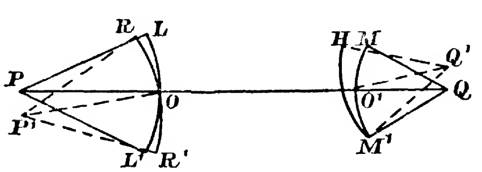


Fig. 102.

width at MM', the angles POP' and QOQ' are $2RL/D$ and MH/d respectively, and are therefore in the inverse ratio of $D:d$. It follows that

$$\frac{PP'}{QQ'} = \frac{d \times PO}{D \times O'Q} \quad \text{.........................(6).}$$

If a square of surface S and sides PP' is formed in a plane at right angles to OO', its image s will be a square with QQ' as sides, hence

$$\frac{S}{s} = \frac{d^2}{D^2} \cdot \frac{PO^2}{O'Q^2}.$$

The solid angle (ω) of the beam entering the first lens is $\pi D^2/4PO^2$, and the solid angle (ω') of the beam converging to Q is $\pi d^2/4O'Q^2$.

Hence the illumination per unit surface of s is

$$\frac{IS\omega}{s} = I \cdot \frac{d^2}{D^2} \frac{PO^2}{O'Q^2} \cdot \frac{\pi D^2}{4PO^2}$$

$$= I\frac{\pi d^2}{4O'Q^2} = I\omega' \quad \text{.........................(7).}$$

Before discussing the last equation, we note two interesting results which have incidentally been obtained in the investigation.

QQ' is inverted as compared with PP' and this must always be the case according to the construction when the limiting ray MQ is the continuation of the ray PL on the same side of the axis, but if the rays have crossed once or an odd number of times between O and O', so that the ray PL becomes the ray $M'Q$, we should have to turn round the ultimate wave-front MM' in the opposite direction in order to equalize the optic lengths of the extreme rays, and the image would then be erect.

The ratio of the angles $QO'Q'$ and POP' is the magnifying power (m) of the arrangement, hence

$$m = \frac{QQ'}{QO'} \div \frac{PP'}{PO},$$

and, by (6),

$$= \frac{D}{d}.$$

In a telescopic system D and d represent the width of the incident and emergent beam respectively, and we have therefore proved the proposition which has already been made use of in Art. 74.

The theorem, defined by equation (7), that the brightness of a luminous surface is determined by the solid angle of the converging pencil which forms the image, is of fundamental importance. We may derive three separate conclusions from it. (1) The apparent brightness of a luminous surface looked at with the naked eye is independent of its distance from the observer. (2) No optical device can increase the apparent brightness of a luminous surface above what it is when the surface is looked at with the naked eye. (3) When looked at through a telescope the brightness of a surface is independent of magnifying power up to a certain limit, and above that limit, the brightness varies inversely as the square of the magnifying power.

The first of these propositions depends on the fact that when looked at with the naked eye, the solid angle on which the brightness depends, is determined solely by the width of the pupil, and the dimensions of the eye; and, independently of casual changes of the pupil, is constant. Hence the brightness of the solar disc is the same when looked at from the furthest or from the nearest planet. The total amount of luminous radiation no doubt diminishes as the distance increases, but the apparent size of the disc diminishes in the same ratio, and hence follows the equality of the amount of light *per unit surface* of the image on the retina. Elementary considerations are sufficient to show that the apparent size of the image of a surface varies inversely as the square of the distance and that illumination is therefore constant, but the second and third of the above propositions are not quite so obvious. Imagine a surface, *e.g.* the moon, looked at through a telescope having an aperture of diameter D. So long as the magnifying power is less than D/p, where p is the diameter of the pupil, the width of the beam entering the eye is p, and the solid angle ω' is the same as if the moon were looked at with the naked eye. The moon would therefore appear to be of exactly the same brightness in the two cases, if there were no loss of light by reflexion and absorption in the optical media of the telescope: *in no case can the moon appear brighter through the instrument.* When the magnifying power (m) is greater than $m' = D/p$, the width of the emergent pencil is $d = D/m$ and the solid angle ω' is reduced in the ratio d^2/p^2 or D^2/p^2m^2. Hence for magnifying powers greater than m', the brightness is reduced into the ratio m'^2/m^2. In observing luminous surfaces, therefore, through a telescope, we may apply magnifying powers up to D/p without loss of brightness except through reflexion and absorption, but we do not make use of the full aperture, and therefore of the full resolving power until the magnifying power has reached that value. Taking the aperture of the pupil to be 3 mm. this would give a magnifying power of $3\frac{1}{3}$ for each centimetre or about nine per inch of aperture. There is, however, an advantage in using somewhat higher magnifying powers, as the outer portions of the crystalline lens do not assist the definition on account of aberration. Most eyes see objects therefore more distinctly when the size of the pupil is reduced to about 2·5 mm. which would give a magnifying power of 4 for each cm. of aperture. With greater magnifying powers, there is no gain in definition and there is loss in brightness. It should be noted that in all cases so far considered the brightness of the image does not in any way depend on the focal length of the lens. It is otherwise when telescopes are used for photographic purposes. The solid angle ω' on which the brightness depends varies in this case with $(D/f)^2$, D being the diameter and f the focal length. A short focus lens of large diameter is therefore of considerable advantage in these cases.

79. Brightness of Stars. The above results apply only so long as the size of the image of a surface is large compared with the size of the diffraction image. Other considerations regulate the brightness of the image of a star. The diameter of the diffraction image of a star has been shown to be inversely proportional to the aperture. When looked at with the naked eye, or through a telescope of low magnifying power, the diameter of the disc is determined by the width of the pupil, and the brightness varies in that case as the amount of light which enters the eye. If the magnifying power is D/p, the amount of light collected by the lens is D^2/p^2 times that collected by the unaided eye. Hence the illumination of the image of a star varies as the square of the effective aperture of the lens, so long as the magnifying power is adjusted so as to be equal to D/p. If less than that, we must imagine the unused portions of the lens to be covered and the aperture reduced to its "effective" portion. When the magnifying power is D/d, d being smaller than p, the linear size of the diffraction image is increased in the ratio p/d, so that the brightness now will vary as D^2d^2/p^4. For star images as well as for finite surfaces there is therefore loss of light without gain in definition, when the magnifying power is increased above a certain value. Astronomers frequently, however, use a higher power than that which according to the above should give the best results. The reason is physiological: Increased size of the diffraction images assists facility of observation, and increases therefore what may be called *physiological* definition, even though there is loss of light and no improvement of *optical* definition.

The increased visibility of stars through telescopes is easily explained. Apart from loss of light due to absorption and reflexion the brightness of the sky is the same whether we use an instrument or not, but the brightness of a star increases with the square of the aperture so that even an opera-glass having an aperture of not quite an inch should increase the light fifty times. The largest telescopes at present in use allow the light which enters the eye from each star to be multiplied by 100,000. It is not surprising then that the number of visible stars becomes rapidly greater as larger apertures are brought into action.

The angular diameters of planets occupy an intermediate position between that of the moon and of the fixed stars. When looked at with the naked eye, the diameter of the image on the retina is less than that of the diffraction disc, but with Venus, Jupiter and Saturn it is only a few times smaller. The use of a telescope having an aperture up to ten times the diameter of the pupil would when applied to these planets be accompanied by an increase of brightness, but after that point is reached, they would behave like bodies of finite surface and an increased aperture would be an advantage because it would allow a higher magnifying power to be applied.

80. Powers of Spectroscopes. A spectroscope may be used for two different purposes. In the majority of cases it serves to examine the radiations of a luminous source by separating them if homogeneous, or giving the distribution of illumination for different wave-lengths, if non-homogeneous. But another not less important function of the spectroscope is to produce homogeneous light. By allowing the spectrum formed by a source of white light to fall on a screen with a narrow slit placed so that only rays very near those of a certain wave-length pass through the slit, we obtain a source of nearly homogeneous radiations. The power of a spectroscope may be defined —either as its power to produce homogeneous light, or as its power to separate two homogeneous radiations of nearly the same wave-length. Both conditions lead to the same mathematical expression, which we now proceed to establish.

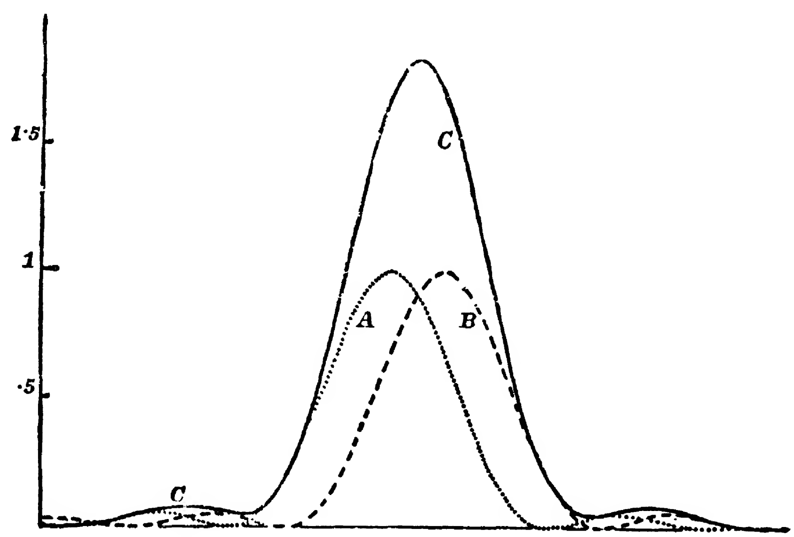


Fig. 103.

The radiations sent out by luminous vapours are often very complicated and sometimes consist of one or more nearly homogeneous radiations lying close together. Consider a source of light sending out waves, the lengths of which, λ_1 and λ_2, differ but little from each other. If the light, after having passed through a slit and been made parallel by a "collimator," falls on a grating, and is then collected by a lens, two images will be formed at the focus. The diffraction image of each is of the same kind as that of a luminous line in a telescope, the object glass of which has been covered by a screen with a rectangular aperture, because the grating itself causes the cross section of the effective beam to be rectangular. If the difference between λ_1 and λ_2 is very small, there is a considerable overlapping, and what the eye perceives is the sum of the illuminations due to each of the two images. In Fig. 103 the curves *A* and *B* show the distribution of intensity of the two separate slit images, while *C* gives the sum of the intensities. The combined curve is so nearly equal to the curve of the image of a single slit that the eye could not realize that the light is made up of two different wave-lengths. The two lines are not in that case "resolved." Fig. 104 gives the combined intensities of the same two lines, when placed three times as far apart, and at such a distance that the maximum illumination of one image falls on the first minimum of

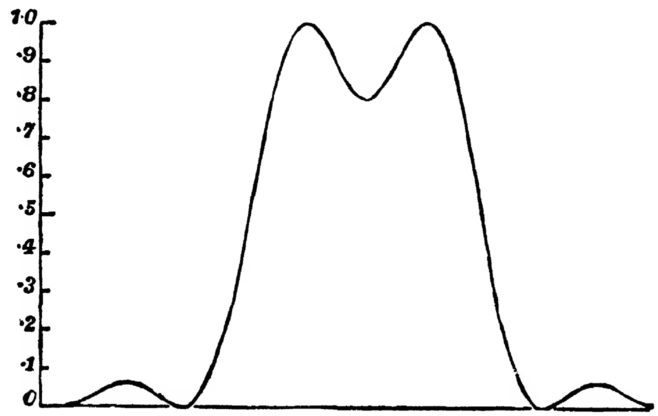


Fig. 104.

the other. The curve shows in this case a decided dip in the middle between two maxima. The intensity at the lowest point is very nearly ·8 of the intensity at the maximum, and the eye clearly perceives that it is not dealing with a homogeneous radiation. The natural interpretation of a distribution of intensity such as that indicated in Fig. 104 is that the radiation consists of two homogeneous radiations having wave-lengths corresponding to the positions of the maxima. The two lines are then said to be "resolved," but it is of course possible, and frequently the case, that the radiation is of a more complicated character. Not until the distance between the two lines is about double that indicated in the figure do they stand altogether clear of each other. According to Art. 60, two wave-lengths λ and λ' have the relative position indicated by Fig. 104, if

$$nN\frac{\lambda' - \lambda}{\lambda} = \pm 1,$$

N being the number of lines on the grating and n the order of the spectrum. In order just to resolve lines with this difference in wave-lengths, $\delta\lambda$ must be such that

$$\frac{\delta\lambda}{\lambda} = \pm \frac{1}{Nn}.$$

The smaller $\delta\lambda$ the more powerful is the instrument for the purpose of separating double lines, and we call as already pointed out Nn the "resolving power" of the spectroscope. There is something arbitrary in this definition, as the dip in intensity necessary to indicate resolution is a physiological phenomenon, and there are other forms of spectroscopic investigation besides that of eye observation. In a photograph or a bolometer, the test of resolution is different. It would therefore have been better not to have called a double line "resolved" until the two images stand so far apart, that no portion of the central band of one overlaps the central band of the other, as this is a condition which applies equally to all methods of observation. This would diminish to one half the at present recognized definition of resolving power. Confusion would result from a change in a universally accepted definition, but it should be understood that if R is the resolving power, a grating spectroscope will completely separate two wave-lengths differing by $\delta\lambda$ only when

$$\frac{\delta\lambda}{\lambda} = \frac{2}{R}.$$

81. Resolving Power of Prisms. It has been proved in Art. 24 that if in a parallel beam of light, two sets of waves are originally superposed, the angle θ between the two beams after passing through a prism is

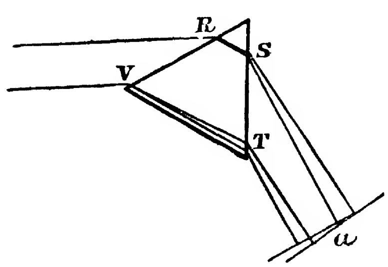


Fig. 105.

$$\theta = (\mu_2 - \mu_1)\frac{VT - RS}{a},$$

where μ_2 and μ_1 are the two refractive indices, and a the width of the beam after emergence. The difference $VT-RS$, for which we write t, is the difference between the paths inside the prism of the extreme rays of the beam. If the prism is placed so that one of the extreme rays just passes by the edge, $RS = 0$ and t will measure the greatest thickness of the dispersive material through which the ray has passed. It is easy to extend the investigation so as to include any number of prisms. If $T = \Sigma t$ measures the difference in aggregate thickness of the material through which the extreme rays have passed,

$$\theta = (\mu_2 - \mu_1)\, T/a \quad\quad (8),$$

the material here being considered the same for all prisms.

This expression leads at once to the resolving power of prism spectroscopes. The beam passing through the prisms having a rectangular cross-section, the angle subtended at the centre of the focussing lens by the half width of the central diffraction band is λ/a (Art. 54), hence with the definition of resolution of the last article, we have for the two wave-lengths at the point of separation, $\theta = \lambda/a$ or

$$\frac{(\mu_2 - \mu_1)}{\lambda} = \frac{1}{T}.$$

The ratio $(\mu_2 - \mu_1)/(\lambda_2 - \lambda_1)$ is the rate of increase of refractive index per rate of increase of wave-length, and may for small differences of wave-length be written $d\mu/d\lambda$. Hence for the resolving power

$$\frac{\delta\lambda}{\lambda} = 1\bigg/\left(T\frac{d\mu}{d\lambda}\right),$$

and

$$R = -T\frac{d\mu}{d\lambda}.$$

The minus sign is necessary, as R is essentially positive and $d\mu/d\lambda$ negative.

This fundamental relation, due to Rayleigh, shows that the resolving power of prism spectroscopes is proportional to the greatest thickness of the dispersive material traversed by the rays (the edges of the prisms being arranged along the path of one of the extreme rays of the beam).

The distinction between the dispersion and the resolving power is a very important one. Confining ourselves to one prism, the dispersion $\theta/(\mu_2 - \mu_1)$ is obtained from (8) and varies inversely as the cross-section of the beam. If a prism be placed in one of the two positions A and B (Fig. 106), the position A gives a greater dispersion than B, in the ratio of $t_1/a_1 : t_2/a_2$, but the resolving powers only vary in the ratio $t_1 : t_2$. The greater dispersion is therefore not accompanied by a correspondingly greater resolving power, the reason being that the narrow beam of A, though giving greater dispersion, gives also a broader diffraction image. The increased dispersion means therefore chiefly increased magnification without increased definition.

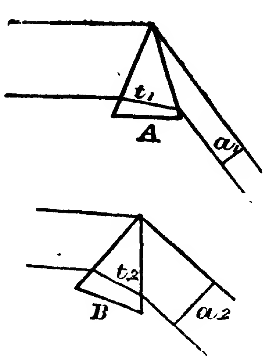


Fig. 106.

With ordinary flint glass $d\mu/d\lambda$ is about 1000 in the neighbourhood of the sodium line, so that one cm. of glass is sufficient to separate the two sodium lines, the difference between their wave-lengths being very nearly equal to the thousandth part of the wave-length of one of them. When the prism is in the position of minimum deviation, the length t is equal to the thickness of the base of the prism. The superiority of gratings over prisms as regards resolving power is shown by the fact that the gratings in common use have about 5600 lines to the centimetre, and if ruled over a distance of 5 cms. the total number of lines would be 28,000. To produce with prisms a spectroscope of resolving power equal to that of the first order spectrum of the grating, would require a thickness of 28 cms. of glass, or say 7 prisms, having a base of 4 cms. each.

It will be noted that the resolving power of prisms depends on the total thickness of glass, and not on the number of prisms, one large prism being as good as several small ones. Thus all the prisms drawn in Fig. 107 would have the same resolving power, though they would show very considerable differences in dispersion.

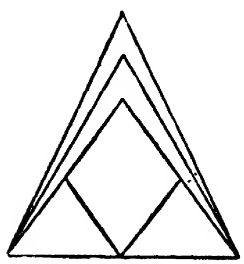


Fig. 107.

82. Resolving power of compound prisms. The only kind of compound prism we need consider here is the one giving direct vision. Two prisms of crown glass, A, A' (Fig. 108), may be cemented to a prism B of flint adjusted so that a ray of definite wave-length falling on the compound prism in the direction of its axis passes out in the same direction. It follows by symmetry that in the centre prism B, the direction of the ray must then also be along the axis. The extreme thicknesses travelled through by the rays are, on one side a thickness t' in flint, and on the other side a thickness $2t$ in crown. The resolving power of such a prism is

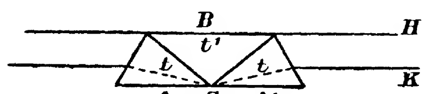


Fig. 108.

$$t'\frac{d\mu'}{d\lambda} - 2t\frac{d\mu}{d\lambda},$$

where $\frac{d\mu'}{d\lambda}$ refers to flint and $\frac{d\mu}{d\lambda}$ to crown. The dispersion of the crown glass is here opposed to that of the flint and the resolving powers of such compound prisms are small taking account of the thickness of glass traversed.

83. Greatest admissible width of source. In considering the effects of interference and diffraction, we had considered the source of light to be either a point or a line, but in the actual experiment, every source has finite dimensions, and as in general it is important to secure as much illumination as possible, these dimensions are increased as much as is consistent with good definition. The limit to which we can increase the dimensions of the source depends to some extent on the object we have in view. When, *e.g.*, we wish to measure accurately a spectroscopic line known to be single, we may use a much wider slit than if we wish to see whether a line is single or double. But even in the last case, there is a limit below which very little is gained by narrowing the source. To determine this limit, we consider the diffraction pattern of the image of a slit. Widening the slit alters the law according to which the intensity of light is distributed in the diffraction image, and it may be seen from an inspection of Fig. 70 that increased width of slit means an increase in intensity which is greater for the weaker portions near the minimum than for the central portion at the maximum. Computation shows however that until the geometrical image of the slit exceeds the eighth part of the width of the central diffraction band, the alteration in the distribution of light is insignificant, so that there is not much advantage in narrowing the slit beyond this limit.

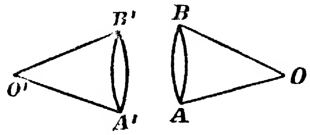


Fig. 109.

Let light passing through the slit at O' (Fig. 109) and the collimating lens $A'B'$ ultimately be focussed by the lens AB. The centre of the image being at the principal focus, the geometrical image of the slit has a width equal to the eighth part of the diffraction band when $BO - AO = \lambda/8$, O being the image of one of the edges of the slit. Fermat's principle at once leads to the conclusion that if O' is the edge of the slit which has its image at O, $A'O' - B'O' = \lambda/8$. The width of the slit is then found by geometrical considerations to be $f\lambda/4D$, where D is the aperture and f the focal length of the collimator lens. Writing ϕ for the angle subtended by the collimator at the slit, the greatest admissible width (d) of slit, for full definition, becomes

$$d = \frac{\lambda}{4\phi}.$$

ϕ is generally about 1/16, so that the width of the slit should not be more than four wave-lengths. When extreme definition is not required, we may, without seriously interfering with the accuracy of the observations, allow a difference in phase of a quarter of a wave-length between the extreme rays. This would double the admissible width of slit. Two spectrum lines placed in the position of Fig. 104 would show with this width of slit a diminution in intensity amounting to 10 % at the lowest point of the intensity curve, instead of 20 % which they give with an indefinitely narrow slit. The resolution would be more difficult, but under favourable circumstances not impossible, as to some extent, the smaller variation in intensity is counterbalanced by increase in brightness. The above condition $A'O' - B'O' = \lambda/8$ may conveniently be expressed in a different form. Let O'' be the other edge of the slit, then by symmetry $A'O' = B'O''$ and hence

$$B'O'' - B'O' = \lambda/8.$$

We may say therefore that for perfect definition the admissible width is determined by the condition that the distances from different points of the source to any one point of the edge limiting the beam shall not exceed one eighth part of the wave-length. In many cases this difference may be increased to one quarter of a wave-length. When put in this form the proposition is of great practical utility. Thus if the bright spot at the centre of the shadow of the circular disc of diameter d is to be observed and f be the distance from the disc of a small opening, through which the light enters, the proposition shows that the greatest admissible linear dimension of the opening is $f\lambda/2D$.

84. Brightness of image in the spectroscope. When dealing with homogeneous light, the investigation of Art. 78 can be applied and we get the maximum illumination when the pupil is filled with light. For a given prism or grating this determines the magnifying power of the observing telescope, at which there is both full resolving power and full illumination. The former is lost by diminishing, while illumination is lost by increasing the magnifying power. Errors are frequently committed, owing to the belief that illumination depends on the ratio of the aperture to the focal length either in the collimator or the telescope of the spectroscope. This is not correct. It is important, however, to consider both the vertical and horizontal aperture of the beam as it leaves the grating or prism. The prism narrows or widens the beam, unless it is in a position of minimum deviation, and with a grating the width depends on the angle at which the spectra are observed. When it is important to magnify, even at the risk of losing light, the spectroscopes have an advantage over telescopes, as by placing the prism out of minimum deviation, in such a way that the beam is narrowed, we enlarge the image in one dimension only, and it is just that lateral magnification which is required. Hence the corresponding loss of light is less than it would be if the enlargement were done by a higher power eye-piece. The so-called half-prisms act in this way, spectroscopes being constructed of considerable magnifying power but comparatively small resolving power by cutting the direct vision compound prisms (Art. 82) into two equal halves at right angles to the axis and using one of the halves only. Light falling on the face which stands perpendicular to the axis enters the prism with a width BC (Fig. 108) and leaves it with a reduced width HK.

When the light is weak, the slit of a spectroscope has often to be widened more than is consistent with full resolving power and this has to be taken into account when designing spectroscopes which are likely to be used to analyse feeble sources. It is convenient to introduce a term which specifies the resolution which an instrument is giving under the actual conditions in which it is used. I therefore define $\lambda/d\lambda$ to be the "purity" of the spectrum if $\delta\lambda$ be the difference between two wave-lengths which just do not encroach upon each other. With indefinitely narrow slits the purity becomes identical with the resolving power, and generally we may write, if P be the purity and R the resolving power: $P = pR$, where p, a function of the slit-width, may be called the purity factor. To establish a connexion between the purity factor and the slit-width we must fix on the separation $\delta\lambda$ which we consider sufficient for resolution. For narrow slits we have already agreed on a test of separation and seen that when two lines are just resolvable the illumination at the point which is half-way between them is ·81 of that which holds for the two maxima. We may then take as a convenient criterion for separation, also in the case of wide slits, that separation which gives the same diminution of light between the two slit images. The way in which the purity and intensity vary with the slit images is shown in Table IX.* The width of the slit is given in units such that it is obtained in centimetres by multiplication with $\lambda f/d$ where f and d are respectively the focal length and diameter of the collimator lens.

TABLE IX.

Width of slit	Purity Factor	Intensity
0 $\times \lambda f/d$	1	0
0·25	·986	·246
0·5	·943	·467
1	·780	·774
1·5	·579	·890
2	·450	·903
3	·311	·931
∞	0	1

The table shows that with a slit width of $\lambda f/4d$ there is only a slight loss of purity, and that width may be considered to be the normal one for sufficiently strong sources of light—a further narrowing would only lead to loss of light without appreciable gain in resolving power. If the slit be widened the illumination increases almost in direct proportion of the slit-width and at first the loss in purity is not serious, but when the width becomes equal to $1·5\lambda f/d$ the intensity is about 11 % short of what it would be with an indefinitely wide slit and there is a loss of 42 % in purity. Beyond that point very little is gained in illumination and much is lost in purity by a further widening of the slit. The general conclusion to be drawn from the table is that for weak sources of light the spectroscope should be designed so as to give about twice the required resolving power and that

* For a more complete Table and treatment of the subject see "The Optics of the Spectroscope." *Astrophysical Journal*, Vol. 21, p. 197 (1905).

the slit may then be opened so as to have a width somewhat less than $2\lambda f/d$. A practical method of opening the slit to that width is based on the consideration that when a parallel beam of light falls on the slit in the direction of the axis of the collimator, the central diffraction image just covers the collimator lens when the width of the slit has the required value.

Although our test of resolution contemplates eye observations, the conclusions we have drawn may be to a great extent applied to cases where either photographic or bolometric methods are used, especially as regards the comparison with each other of the powers of different instruments. When objective methods are used, greater illumination may be secured by reducing the focal length of the lens which forms the ultimate image, but on the other hand there is generally a limit below which the linear scale of the final image cannot be reduced so that the best conditions for subjective and objective methods of observations do not differ so much as might at first sight appear.

Full resolving power is only obtained if the collimator lens is completely filled with light. Hence when the slit S is wide, and the source of light (L) is narrow, it is necessary to interpose a lens (A) as shown in Fig. 110; the angular aperture of the lens A as seen from the slit need not, in this case, be larger than the angular aperture of the collimator lens. If the diameter of the lens A is increased without changing its focal length, the image of L on the slit plate becomes brighter, but the increase in light is caused by rays which do not pass through the collimator lens at all, and are therefore useless. When the slit is made so narrow that full or nearly full resolving power is obtained, diffraction will cause the light inside the collimator to spread, so that in the arrangement as drawn in the figure a good deal of light is lost. To make up for this, the diameter of A should be increased, when some of the light which with a wide slit falls on the sides of the collimator tube, will be diverted so as to pass through the collimator. In the observation of star spectra, the telescope lens performs the function of the condensing lens A, and its aperture being fixed, there is necessarily a not inconsiderable loss of light due to diffraction, when full resolving power is obtained. Diminishing the focal length of the collimator does not help here, because this would have to be accompanied by a narrowing of the slit, if the resolving power is to be maintained. Hence, as in other cases where total light is a consideration, it becomes necessary to build the spectroscopes so as to have a resolving power twice that actually required, and to open the slit to a width of about $2\lambda f/d$.

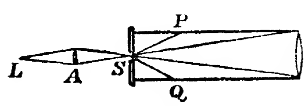


Fig. 110.

It should be noted that if a star image were perfectly steady and undisturbed by irregular atmospheric refraction, a star spectroscope should give full resolving power without any slit at all. Indeed in this case, the slit could only deteriorate, but never improve the definition. The tremors of the star images, due to atmospheric fluctuations, are however sufficiently serious to render a slit desirable, when high resolving powers are required.

The above treatment of the subject is based on the consideration of spectra of bright lines, and cannot without modification be applied to the absorption phenomena exhibited in the spectra of sun and stars. It would lead too far to enter into this part of the subject here, but one example of the difference presented by emission and absorption spectra may be pointed out. A perfectly homogeneous radiation could never appear as a dark line in an absorption spectrum, for the reason that an indefinitely narrow gap between two bright surfaces could not be detected by any instrument of finite resolving power.

85. Aberrations. To obtain the maximum concentration of light at a point the wave-fronts should be perfect spheres with that point as centre, but owing to defects in the working of the surfaces, or insufficient homogeneity of the material, perfection is never attained. From the point of view of the wave-theory of light, the so-called optical "aberrations" are dependent on the deviations of the wave-fronts from the ideal spherical shape. The amount of deviation compatible with sensibly perfect definition has been discussed by Lord Rayleigh*, who finds that if the discrepancy in phase at the focus between the extreme and central rays of the wave-front does not exceed a quarter of a wave-length, the image does not suffer appreciably. At that limit, a rapid deterioration of definition begins. Lord Rayleigh also gives the following important applications of this result. If in a telescope supposed to be horizontal, there is a difference in temperature between the stratum of air along the top and that of the rest of the tube, the wave-fronts are distorted along the top of the tube. The final error of optical length of the extreme rays is $l\delta\mu$, where l is the length of the tube and $\delta\mu$ the alteration in refractive index. At ordinary temperatures $\delta\mu$ is connected with δt the change in temperature, by the approximate relation

$$\delta\mu = -1{\cdot}1\,\delta t \times 10^{-6}.$$

If the error in optical length is a quarter of a wave-length,

$$\tfrac{1}{4}\lambda = 1{\cdot}1\,l\delta t \times 10^{-6},$$

$$\therefore\ l\delta t = 12 \text{ if } \lambda = 5{\cdot}3 \times 10^{-5}.$$

Thus a change of temperature of 1° becomes appreciable when the length through which the temperature difference extends is 12 centi-

* *Collected Works*, Vol. I. p. 428.

metres. In a telescope tube 12 metres long, the average temperature of the air through which the different rays pass should not differ by more than 0·01 degrees.

As a second example, also given by Lord Rayleigh, we may take the accuracy which is required in the working of optical surfaces. If AC is the optical surface and if through imperfections in the working any portion of it is raised so as to occupy the position DF, the error in optical length is (Fig. 111)

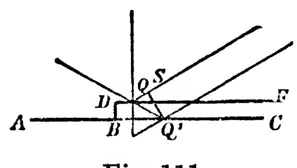


Fig. 111.

$$QQ' - QS = 2BD\cos\phi,$$

where ϕ is the angle of incidence and S the foot of the perpendicular from Q' to the ray reflected from Q. Hence the deviation BD from the plane AC should not, over any considerable portion of the surface, exceed $\frac{1}{8}\lambda\sec\phi$, or for normal incidence, one-eighth of the wave-length.

Our result applies to the case where no change of focus is allowed in the observing telescope, but aberrations, in the sense here introduced, may often be corrected for by such change of focus as *e.g.* when a surface intended to be plane is slightly concave or convex.

Students may, as an example, work out the greatest admissible width compatible with perfect definition of a spherical mirror, when rays parallel to the axis fall on it.

86. The formation of images without reflexion and refraction. Pin-hole Photography. An elementary experiment in Optics consists in showing the rectilinear propagation of light by projecting an image on a screen, the image being formed by rays which have passed through a narrow aperture. Lord Rayleigh has shown that for small apertures such an opening acts as well as a lens, and the discussion of the matter is here given in his own words :—

"The function of a lens in forming an image is to compensate by its variable thickness the differences in phase which would otherwise exist between secondary waves arriving at the focal point from various parts of the aperture. If we suppose the diameter of the lens ($2r$) to be given, and its focal length f gradually to increase, these differences of phase at the image of an infinitely distant luminous point diminish without limit. When f attains a certain value, say f_1, the extreme error of phase to be compensated falls to $\frac{1}{4}\lambda$. Now, as I have shown on a previous occasion*, an extreme error of phase amounting to $\frac{1}{4}\lambda$, or less, produces no appreciable deterioration in the definition ; so that from this point onwards the lens is useless, as only improving an image already sensibly as perfect as the aperture admits

* *Phil. Mag.* Nov. 1879.

of. Throughout the operation of increasing the focal length, the resolving-power of the instrument, which depends only upon the aperture, remains unchanged; and we thus arrive at the rather startling conclusion that a telescope of any degree of resolving-power might be constructed without an object-glass, if only there were no limit to the admissible focal length. This last priviso, however, as we shall see, takes away almost all practical importance from the proposition.

"To get an idea of the magnitudes of the quantities involved, let us take the case of an aperture of $\frac{1}{5}$ inch (inch $= 2{\cdot}54$ cms.), about that of the pupil of the eye. The distance f_1, which the actual focal length must exceed, is given by

$$\sqrt{\{f_1^2 + r^2\}} - f_1 = \tfrac{1}{4}\lambda,$$

so that

$$f_1 = 2r^2/\lambda.$$

Thus, if $\lambda = \frac{1}{40,000}$, $r = \frac{1}{10}$, $f_1 = 800$.

"The image of the sun thrown on a screen at a distance exceeding 66 feet, through a hole $\frac{1}{5}$ inch in diameter, is therefore at least as well defined as that seen direct. In practice it would be better defined, as the direct image is far from perfect. If the image on the screen be regarded from a distance f_1, it will appear of its natural angular magnitude. Seen from a distance less than f_1 it will appear magnified. Inasmuch as the arrangement affords a view of the sun with full definition and with an increased apparent magnitude, the name of a telescope can hardly be denied to it.

"As the minimum focal length increases with the square of the aperture, a quite impracticable distance would be required to rival the resolving-power of a modern telescope. Even for an aperture of four inches f_1 would be five miles."

Returning to the subject in a later paper, Lord Rayleigh discusses its application to the so-called pin-hole photography, in which the lens of a camera is simply replaced by a narrow aperture. If this aperture is too small, the image loses in definition owing to the spreading out of the waves, and on the other hand it is clear that no image can be formed, when the aperture is large. There must therefore be one particular size of the opening which gives the best result. The original paper* should be consulted, in which the question is treated both theoretically and experimentally. The best result in general is found, when the aperture as seen from the image includes about nine-tenths of the first Fresnel zone, so that if a is the distance of the object, b that of the image from the screen and r the radius of the opening,

$$r^2 \frac{a+b}{ab} = {\cdot}9\lambda.$$

* *Collected Works*, Vol. III. p. 429.

86 A. Photometry. The methods and appliances used in comparing the intensities of two sources of white or nearly white light are described in experimental text books and lie outside the range covered by this treatise. When the radiations which are to be compared with each other differ appreciably in colour, difficulties arise in the selection of a common standard unless the thermal effects are sufficiently great to allow the intensities to be measured directly. Nicholson and Merton have shown how these difficulties can be overcome. When the intensities of two nearly homogeneous radiations differ only slightly in wave-length, the method first used by Merton* is sufficient. It consists in transmitting the light before it enters the spectroscope through a neutral-tinted wedge of glass mounted so that the lines of equal thickness of the wedge are at right angles to the length of the slit. The spectrum of a luminous gas then appears to consist of lines bright at the points corresponding to the thin end of the wedge and gradually falling off in intensity. If I_1 be the intensity of a certain line it is reduced to $I = I_1 e^{-\kappa x_1}$ in passing through a thickness x, of the wedge. Similarly for a second line $I = I_2 a^{-\kappa x_2}$. If two of the resulting intensities I are the same in both cases: $\log (I_2/I_1) = \kappa (x_2 - x_1)$. Hence the relative intensities of the two radiations can be determined if κ be known and $x_2 - x_1$ determined by the observation.

Before the method can be applied to lines in different parts of the spectrum several limitations must be removed. In the first place the absorbing power of the wedge depends on the wave-length. To determine κ for a line having a given wave-length, equal exposures are taken with intensities of the incident light reduced in a known proportion†. Let l_1 and l_2 be the measured lengths of the lines, the incident light being I and aI respectively. If α be the angle of the wedge, m the magnifying power, and I_c denote the intensity at which a line is just photographically visible, we have:

$$\log aI/I_c = \kappa l_1 m^{-1} \tan \alpha; \quad \log I/I_c = \kappa l_2 m^{-1} \tan \alpha.$$

Hence
$$\log a = \kappa (l_1 - l_2) m^{-1} \tan \alpha.$$

The last equation determines κ.

It is more difficult to correct for the differences in the photographic sensitiveness in the different parts of the spectrum. The device used for the purpose by Nicholson and Merton was to take a photographic record of the continuous spectrum of the positive crater of the electric arc. This has a definite temperature not differing much from 3650°. Assuming that the radiation corresponds to that of a perfectly black body, the varying sensitiveness of different photo-

* *Proc. Roy. Soc.* A, Vol. XCII. p. 322.

† *Phil. Trans.* A, Vol. CCXVII. p. 237.

graphic plates may be determined by comparing the record of the photographic intensity of the spectrum obtained with the theoretical radiation of a black body.

86 B. Measurements of the Diameters of Stars. We may conclude this chapter with a short account of the remarkable results which Michelson has obtained by applying interference methods to increase the resolving powers of telescopes. We have seen (Art. 73) that the image of a luminous point formed by an ordinary lens consists of a central disc surrounded by alternating bright and dark rings and that the radius of the disc, on which the resolving power depends, is $1{\cdot}22\lambda f/2R$, where f is the focal length and $2R$ the diameter of the lens. If the lens be covered with a diaphragm having a horizontal aperture a and a vertical height b, which we shall take to be small, the image of a point consists of vertical diffraction bands of diminishing intensity as their distance from the central band increases. Equation 5, Art. 76, shows that the distance of the first dark band from the centre, obtained by giving to α the value π, is $\lambda f/a$. Comparing rectangular and circular apertures, the width of the opening of one being equal to the diameter of the other ($a = 2R$), we find that the rectangular aperture has a slightly greater resolving power. This may be still further increased by blocking out the central portion of the lens, and Michelson, pushing the advantage to its limits, covered the lens almost completely, leaving only an aperture of small dimensions at each of the two ends of a diameter. A lens covered in this way cannot of course be used for ordinary purposes of seeing, the resolving power being confined to one direction only.

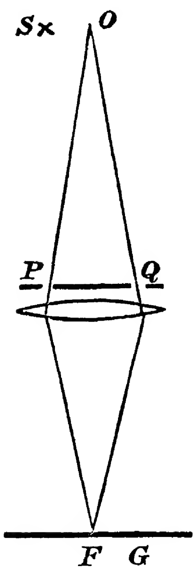


Fig. 111 a.

In describing the distribution of light in the image of a source of light formed in this manner we shall imagine the axis of the telescope to be in a horizontal position and the line joining the two openings to be also horizontal.

In Fig. 111 a let O be a luminous point. If F be the geometrical image of O formed by a lens covered by a diaphragm, having small apertures at P and Q, an interference pattern will appear in the focal plane of the lens with F as centre. The resultant amplitude is that found in Art. 30 for two luminous sources vibrating in the same phase. At a distance x from F, the amplitude is proportional to $\cos \kappa x$, where $\kappa = \pi a/f\lambda$. The first minimum occurs when $x = f\lambda/2a$. It is convenient to use a unit of amplitude such that the average illumination in the plane considered is unity. The illumination due to the interfering sources is then $2\cos^2 \kappa x$.

If a second luminous point be placed at S, its geometrical image being G, the previously adopted test of resolving powers leads to the conclusion that the two sources of light will be just resolved when $FG = f\lambda/2a$. The resolving power would therefore be just double that found for a rectangular aperture having a horizontal width a and nearly two and a half times as great as with a circular aperture. In comparing the efficiency of the two arrangements we must take account of an important difference in the manner in which the resolution of two closely adjacent sources of light is tested. In ordinary telescopic vision we are guided by an illumination curve such as that represented in Fig. 104 and recognize the resolution by a slight diminution of the luminosity between two not very well defined maxima of light. The interference method gives us a much more delicate criterion, the resolution depending on the disappearance of a set of fringes. If ξ be the distance between F and G the illumination of the screen due to the combined sources is

$$I = 2\cos^2 \kappa x + 2\cos^2 \kappa (x + \xi) = 2 + \cos 2\kappa x + \cos 2\kappa (x + \xi);$$

the illumination is therefore uniform and the diffraction pattern disappears when $\kappa\xi$ is an odd multiple of a right angle. There is of course considerable loss of light, only a small part of the lens being used, but this is more than counterbalanced by the greater facility of observation and measurement. Michelson judges that with equal apertures the interference method increases the resolving power about ten times.

The most important application of the interference method is that which leads to the measurement of the angular dimensions of an object so far away that even the most powerful telescopes cannot discriminate between it and a source of light concentrated in a point.

Let the vertical lines in Fig. 111 b represent the maxima of light in the focal plane of a telescope, when the light from a point-source is

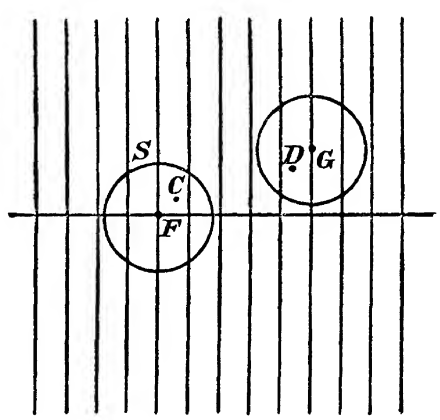


Fig. 111 b.

transmitted as in the previous case through two small apertures at a distance a from each other. Each point of the luminous object, which we may take to be a star, produces identical patterns slightly displaced relatively to each other, and we must find the resultant illumination due to each surface element of the surface. Let S be the geometrical image of the luminous source having F as centre. If G be the point at which the illumination is required, and we take points D and C such that $DG = FC$, it is clear that the luminosity at G due to the interference pattern having its centre at C is the same as the effect of the luminosity at D belonging to the pattern having its centre at F. If this reasoning be applied to all points C of the geometrical image of the star it follows that the total illumination at G is equal to the average illumination in the central interference pattern within a circle having G as centre and a radius ρ equal to that of the geometrical image of the star. Hence

$$I = (\pi\rho^2)^{-1} \iint K \cos^2 \kappa x \, dx \, dy,$$

where K is a quantity depending on the variations in the light emitted from different parts of the surface of the star. For the sake of simplicity we shall take the star disc to be circular and K to be constant and equal to unity.

If we put $x = x_0 + \xi$, where ξ is the horizontal co-ordinate measured from G and x_0 the horizontal distance between F and G, we obtain for the measure of the illumination

$$I = (\pi\rho^2)^{-1} \int_{-\rho}^{+\rho} \int_{-y_1}^{+y_1} \cos^2 \kappa (x_0 + \xi) \, d\xi \, dy,$$

where $$y = \sqrt{\rho^2 - \xi^2}.$$

Performing the integration with respect to y, the illumination becomes

$$(\pi\rho^2)^{-1} \int_{-\rho}^{+\rho} \sqrt{\rho^2 - \xi^2} \{1 + \cos 2\kappa (x_0 + \xi)\} \, d\xi$$

$$= (\pi\rho^2)^{-1} \int_{-\rho}^{+\rho} \sqrt{\rho^2 - \xi^2} (1 + \cos 2\kappa x_0 \cos 2\kappa\xi) \, d\xi;$$

the term omitted on the right-hand side does not affect the result, because it has opposite signs for equal positive and negative values of ξ. The integral $\int_{-\rho}^{+\rho} \sqrt{\rho^2 - \xi^2} \, d\xi$ represents half the area of the circular disc, so that we may write for the illumination at a distance x from the centre of the geometrical image,

$$I = 1 + 2 (\pi\rho^2)^{-1} \cos 2\kappa x \int_{-\rho}^{+\rho} (\sqrt{\rho^2 - \xi^2}) \cos 2\kappa\xi \, d\xi.$$

The integral can now be expressed in terms of a certain function called the Bessel function of the first kind and generally denoted by J_1. When expressed in a series of ascending powers the function takes the form

$$J_1(z) = \frac{z}{2} - \frac{z^3}{2^2 . 4} + \frac{z^5}{2^2 . 4^2 . 6} - \frac{z^7}{2^2 . 4^2 . 6^2 . 8} + \dots .$$

The integral occurring in the equation for I is $\pi\rho\,(2\kappa)^{-1} J_1(2\kappa\rho)$, and the illumination at a point x becomes

$$I = 1 + (\kappa\rho)^{-1} (\cos 2\kappa x) J_1(2\kappa\rho).$$

When ρ is very small so that only the first term of the series needs to be taken into account the expression for the illumination reduces to

$$I = (1 + \cos 2\kappa x) = 2 \cos^2 \kappa x,$$

which, as it should be, is equal to that found for a single luminous point.

The function J_1 fluctuates between positive and negative values and is the same as that which determines the distribution of light of a point source focussed by a lens having a circular aperture. When J_1 is zero, the illumination is uniform and equal to unity. The numerical values of z, etc., for which the function J_1 vanishes, are obtained from the numbers given in the column headed m of Table VII, Art. 73, by multiplication with π. We are now in a position to find a connexion between the dimensions of the dark rings surrounding the image of a luminous point and the distance (a) between the two openings at which the diffraction bands vanish. This distance is given by $z_n = 2\kappa\rho$, if z_n is one of the numbers for which the function vanishes, or substituting the value of κ, we derive for the angular radius of the star

$$\rho = f\lambda z_1/2\pi a.$$

The radii of the dark rings as found by Airy and Lommel are*

$$r = f\lambda z_1/2\pi R,$$

where r is the radius of the circular opening. Comparing the two last equations we derive the following interesting proposition:

The resolving power of a lens of radius R has the same resolving power as Michelson's apparatus for measuring the diameters of stars, having mirrors which are a distance R apart.

It will be noticed that the distance between the two openings is equivalent to the *radius*—not to the *diameter*—of the lens; this illustrates the greater power of the interference method. The observations

* Rayleigh, *Collected Works*, Vol. III. p. 88.

are conducted by altering the interval between the two mirrors symmetrically so as to keep them at equal distances from the centre.

A further and most important increase of resolving power has been obtained by Michelson by lengthening the base line from the ends of which a star is observed. In the arrangement that has been described this base line is the distance between the two openings covering the lens or mirror of the telescope. It was a brilliant idea to increase it by receiving the light on two mirrors that are further apart, and bringing it into the telescope with the help of two additional mirrors. The diagram (Fig. 111 *c*) will illustrate the disposition of the mirrors and the passage

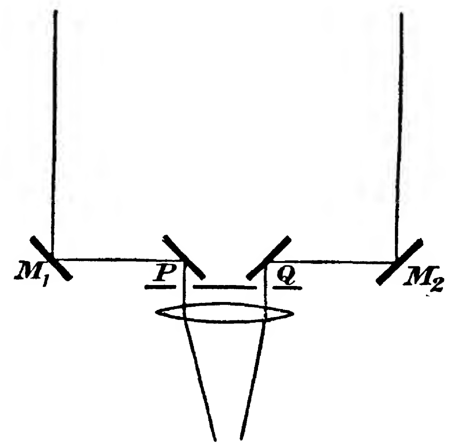


Fig. 111 *c*.

of the light. The distance between P and Q, the two openings in front of the lens, determines the interference pattern and remains constant during the observations, being unaffected by the positions of M_1 and M_2. On the other hand, the distance between these two latter mirrors determines the length we have designated by ρ, and which no longer is the radius of the geometrical image of the star in the telescope, but that which a lens of diameter M_1M_2 would give. This may be demonstrated by a method depending on Fermat's principle, which we have already used on several occasions. If by any optical arrangement a beam of light AB (Fig. 111 *d*) be reduced in width so as to become equal to ab, and AB be turned through a small angle, then ab will be turned through an angle greater in the ratio AB/ab. For if the new wave-front be AC, the optical length from B to b is reduced by BC. In order to keep the optical length from C to c equal to that from A to a, which is the condition that ac shall be the new wave-front, it is necessary that $bc = BC$; the angle bc/ab is in that case greater than BC/AB in the ratio AB/ab. It is assumed here that the media in which the wave-fronts AB and ab lie have the same optical properties. In the case now under consideration AB and ab are not two wave-fronts, but they are points on two rays emanating from the same luminous point and such that the phases at A and B are the same. The reasoning therefore holds good, and our previous equations may be retained if a denotes the distance M_1M_2.

In the first observations made with the arrangement illustrated in Fig. 111 *a*, a reflecting telescope having a parabolic mirror of 100 inches aperture was used, and the distance between the openings P and Q was varied until the diffraction bands disappeared. The distance between

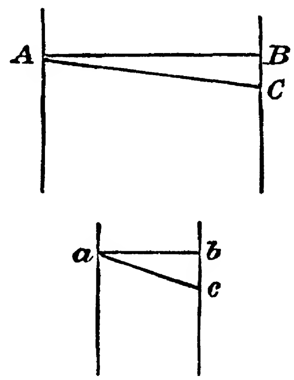


Fig. 111 *d*.

close double stars could thus be measured. When the second arrangement (Fig. 111 *d*) was introduced the distance between the openings P and Q was not altered during the observations, but was reduced to 45 inches so as to give the interference pattern a spacing of ·02 mm. easily visible with a magnification of 1600. The mirrors M_1, M_2 were attached to a beam which allowed a separation of 20 feet, so that the efficiency of the appliance was more than doubled as compared with the possible separation of 100 inches in the previous arrangement. On a first trial it was found that with the mirrors 121 inches apart the fringes of α Orionis were invisible. It is estimated from this that the angular diameter of the star is about 0″·047, and taking account of the measured parallax of ·0018, its linear diameter may be calculated as being 240 million miles, or slightly less than the orbit of Mars. Though it may be observed that the interference pattern is not the same for different wave-lengths, this does not seem to have interfered seriously with the observations, and in the calculations the effective wave-length for α Orionis was assumed to be 5700Å.

86 c. Historical. John William Strutt, Third Baron Rayleigh (born Nov. 12, 1842; died July 1919). The late Lord Rayleigh's contributions to Optics were numerous and important. His name occurs almost in every section of this book, which is dedicated to him, but it may be more specially associated with this chapter, because he introduced clarity and precision into the theory of optical instruments. The erroneous ideas on the resolving powers of prisms and gratings prevalent at the time were finally disposed of by the simple reasoning through which he established the criterion governing the separation of different wave-lengths in Optics as well as in Acoustics. His papers on the nature of light are equally important, and in particular his remark that every homogeneous radiation is necessarily polarized may be quoted because it has not always received the attention it deserves.

CHAPTER VIII.

THE PROPAGATION OF LIGHT IN CRYSTALLINE MEDIA.

87. The Ellipsoid of Plane Wave Propagation. Ellipsoid of Elasticity. It has been shown in Chapter II., Art. 12, that the velocity of propagation of a distortional wave in an isotropic medium is $\sqrt{n/\rho}$, where n is the resistance to distortion. If the medium is not isotropic, the coefficient n will depend on the direction of the displacement. In that case, a plane wave may be propagated with different velocities according to the direction of the vibration. Fixing our mind on a wave-front parallel to a given plane, there must always be one direction for which n has a maximum value n_1, and one for which it has a minimum value n_2. There is therefore a maximum and minimum velocity of propagation $\sqrt{n_1/\rho}$ and $\sqrt{n_2/\rho}$ respectively for every plane with two directions of vibration corresponding to these two velocities. If the displacement is in neither of these two directions, it might seem at first sight that the wave would be propagated with some intermediate velocity. This is not, however, found to be the case either analytically or by experiment. What happens is that the wave splits into two wave-fronts proceeding with the velocities $\sqrt{n_1/\rho}$ and $\sqrt{n_2/\rho}$. If we change the direction of the plane, the two velocities in general change also. It may be proved that the two directions of displacement corresponding to the minimum and maximum coefficients of distortion, are always at right angles to each other. If the direction of displacement be confined to one or other of these two directions, a plane wave may be propagated as a single plane wave. But in the general case, the displacement must be decomposed into its components in the two directions for which a single plane wave propagation is possible. The following construction connecting the velocities of the plane wave propagation in different directions, though originally suggested by theoretical considerations, should at the present stage be considered simply as a representation of experimental facts.

Take some fixed point O (Fig. 112) in the crystalline medium and imagine planes drawn through the point. In each plane take two lines OP_1 and OQ_1 which coincide with the two possible directions of vibration. If v_1 be the velocity of the plane wave for a direction of vibration OP_1 and v_2 for a direction OQ_1, take the points P_1, P_2, Q_1, Q_2, so that $OP_1 = OP_2 = V/v_1$; $OQ_1 = OQ_2 = V/v_2$, V being the velocity of the wave propagation in vacuo*. If the plane through O be altered in direction and the points P and Q marked off for each, it is found that these points lie on an ellipsoid, which may be called the ellipsoid of plane wave propagation. It is also found that the points P and Q lie at the ends of the semiaxes of the central sections of this ellipsoid. If the ellipsoid is given, we may therefore find the direction of vibration and the corresponding velocities of waves parallel to any given plane, by drawing the central section which is parallel to that plane. The semiaxes of the ellipse in which the section cuts the ellipsoid give the directions of vibration, and the velocities are inversely proportional to the semiaxes.

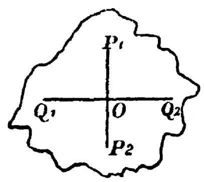


Fig. 112.

Let the equation of the ellipsoid be

$$a^2x^2 + b^2y^2 + c^2z^2 = V^2 \quad\text{.......................(1)},$$

the quantities a, b, c, being in descending order. To simplify the equation, take the unit of time to be the time which a wave in vacuo takes to traverse unit distance, so that we may write $V = 1$. For $x = 0$, we obtain the intersection of the ellipsoid with the plane of yz, which is an ellipse having $1/b$ and $1/c$ as semiaxes. Hence a wave-front may be propagated in the direction of the axis of x either with a velocity b or with a velocity c, the direction of vibration in the former case being the axis of y, and in the latter the axis of z. Similarly a and c are the velocities of propagation for a wave-front parallel to the plane xz, and a, b the velocities for a plane parallel to xy. The velocities a, b and c are called the principal velocities. Fig. 113 illustrates how a plane wave separates into two, the directions of vibration in the two being at right angles to each other.

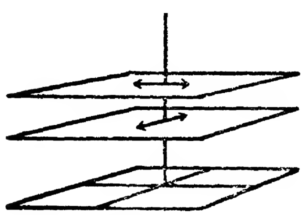


Fig. 113.

We proceed to calculate, in accordance with the general indications given above, the velocities of propagation for any wave-front inclined to the principal axis of the ellipsoid (1). It is required to find the

* We might also take $OP_1 = V/v_2$ and $OQ_1 = V/v_1$ and fit the observed phenomena equally well, but it is convenient to make our choice at once so as to fit in with the view that the direction of vibration is at right angles to the plane of polarization.

direction and magnitude of the principal axes OP and OQ of the intersection of the ellipsoid (1) by a central plane which is defined by the direction cosines l, m, n, of its normal ON, Fig. 114.

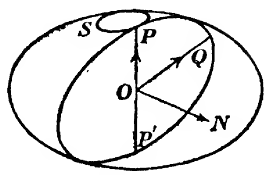


Fig. 114.

Let the length of the semiaxis OP be ρ and construct a sphere of radius ρ having its centre at O. The sphere will pass through the point P and through a corresponding point P' at the opposite end of the same diameter, but it cannot have any other point in common with the ellipse, because the ellipse can only have one diameter equal in length to one of its axes. The sphere intersects the ellipsoid in a curve PS and a corresponding one through P'. These curves, which belong to the class called sphere-conics, must touch the ellipse at P and P'. The same reasoning holds good if all dimensions are diminished in the same ratio. If therefore a cone be constructed having O as vertex and passing through the sphere-conic PS, this cone cannot intersect but must touch the plane of the section OPQ along the diameter PP'.

The equation of the cone is obtained by combining the ellipsoid

$$a^2x^2 + b^2y^2 + c^2z^2 = 1,$$

and the sphere

$$x^2 + y^2 + z^2 = \rho^2,$$

in such a way as to represent a cone passing through the origin. The resulting equation must therefore be homogeneous in x, y. Multiplying the first by ρ^2 and subtracting, we obtain

$$(a^2\rho^2 - 1)\,x^2 + (b^2\rho^2 - 1)\,y^2 + (c^2\rho^2 - 1)\,z^2 = 0,$$

and as the velocity of the plane wave vibrating along OP is $1/\rho$,

$$(v^2 - a^2)\,x^2 + (v^2 - b^2)\,y^2 + (v^2 - c^2)\,z^2 = 0 \quad \text{............}(2).$$

We now introduce the condition that the direction cosines of the normal to the section OPQ coincide with the direction cosines of the plane which is tangent to the cone, the line OP being the line of contact. If x, y, z be now the coordinates of P we obtain in the usual way

$$\left.\begin{aligned} Dl &= (v^2 - a^2)\,x \\ Dm &= (v^2 - b^2)\,y \\ Dn &= (v^2 - c^2)\,z \end{aligned}\right\} \quad \text{..........................}(3),$$

and by substitution in (2)

$$\frac{l^2}{(v^2 - a^2)} + \frac{m^2}{(v^2 - b^2)} + \frac{n^2}{(v^2 - c^2)} = 0 \quad \text{...............}(4).$$

This important equation is of the second degree in v^2 and has two positive roots. There are therefore two possible values of v which satisfy the equation. By getting rid of the denominator and substituting $m^2 = 1 - (l^2 + n^2)$, the equation may be written

$$v^4 - [(a^2 + c^2) - (a^2 - b^2) l^2 - (b^2 - c^2) n^2] v^2 + [a^2c^2 - (a^2 - b^2) c^2 l^2 - (b^2 - c^2) a^2 n^2] = 0 \ldots (4a).$$

Having found v^2 we may find the direction cosines of the line of vibration OP for a given wave-front by substituting in (3) for x, y, z their equivalents $\rho \cos \alpha$, $\rho \cos \beta$, $\rho \cos \gamma$. We have then, with a different meaning of D,

$$\left.\begin{aligned} \frac{Dl}{v^2 - a^2} &= \alpha \\ \frac{Dm}{v^2 - b^2} &= \beta \\ \frac{Dn}{v^2 - c^2} &= \gamma \end{aligned}\right\} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(4b),$$

where D is determined by the relation $\alpha^2 + \beta^2 + \gamma^2 = 1$, which gives

$$\frac{1}{D^2} = \frac{l^2}{(v^2 - a^2)^2} + \frac{m^2}{(v^2 - b^2)^2} + \frac{n^2}{(v^2 - c^2)^2}.$$

With the unit of time adopted, the velocity of light in vacuo is unity, and $1/a$, $1/b$, $1/c$, the reciprocals of the principal velocities, measure quantities which in an isotropic medium would correspond to the refractive index. These quantities are therefore called the principal refractive indices. Denoting them by μ_1, μ_2, μ_3, we may write the equation of the ellipsoid (1) in the form

$$\frac{x^2}{\mu_1^2} + \frac{y^2}{\mu_2^2} + \frac{z^2}{\mu_3^2} = 1.$$

The coefficients of elasticity, which measure the resistance to distortion in the principal planes, are proportional to a^2, b^2, c^2 respectively, so that these constants are intimately connected with the elastic properties of the medium. The ellipsoid (1) has therefore been called the ellipsoid of elasticity (see also Art. 104). In a homogeneous medium, $\mu_1 = \mu_2 = \mu_3$, and the ellipsoid of elasticity becomes a sphere, having a radius numerically equal to the refractive index.

88. The Optic Axes. Every ellipsoid has two circular sections passing through that principal axis which is neither the largest nor the smallest. It follows that there are two directions in which a plane wave-front has only a single velocity. These two directions are called the "*optic axes.*" The radius of the circular section is $1/b$, and putting $\rho = 1/b$ in the equation (2) of the cone, it reduces to

$$(a^2 - b^2) x^2 + (c^2 - b^2) y^2 = 0.$$

This is the equation of the two planes which contain the two circular sections. The two directions of single wave velocities are the normals to these planes, so that

$$l_1 = \pm\sqrt{\frac{a^2-b^2}{a^2-c^2}};\quad m_1 = 0;\quad n_1 = \pm\sqrt{\frac{b^2-c^2}{a^2-c^2}} \quad\text{.........(5)}$$

are the direction cosines of the optic axes.

When a wave is propagated in the direction of one of the optic axes, the direction of vibration may be anywhere in the plane, as in the circular section, any direction may be considered to be an axis.

89. Uniaxal and Biaxal Crystals. In general a crystal has two optic axes and is then called "biaxal." If two of the principal axes are equal to each other, there is only one optic axis, which is the axis of x if $b=c$, and the axis of z when $a=b$. The crystal is then said to be a "uniaxal" crystal.

The ellipsoid of elasticity for uniaxal crystals when $a=b$, is the spheroid

$$a^2(x^2+y^2)+c^2z^2=1$$

and the equation (4) for determining the velocities of plane wave propagation becomes, writing θ for the angle between the optic axis and the normal,

$$\frac{\sin^2\theta}{a^2-v^2}+\frac{\cos^2\theta}{c^2-v^2}=0,$$

or

$$v^2=c^2\sin^2\theta+a^2\cos^2\theta.$$

Hence the velocity depends only on the angle which the normal to the wave-front makes with the axis of revolution of the spheroid.

90. Wave-Surface. The passage of waves through crystalline media is completely determined by the equation we have obtained for the propagation of plane waves, but it is often convenient to base our investigations on a surface which is the locus of equal optical distances measured from a point as centre. Such a surface, according to the definition of Art. 18, is called a "Wave-Surface." Its relation to the optical distance between parallel wave-fronts as deduced in the last paragraph may best be seen by applying Huygens' principle. Let a plane wave (Fig. 115), WF, be propagated upwards and with points P_1, P_2, P_3, etc. as centres construct the surfaces of equal optical distance, corresponding to unit time, *i.e.* the wave-surfaces ST and $S'T'$. The furthest distance to which the wave-front WF can have gone in the time is the tangent plane $W'F'$ to all the wave-surfaces, and by Huygens' principle, as explained in Art. 16, this plane will actually be the position of the wave-front after unit time. The lines which join the centres of disturbance P_1, P_2, P_3, etc. with the

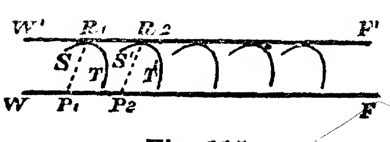


Fig. 115.

points of contact R_1, R_2, R_3, etc. of the wave-surfaces and wave-front, are the lines of shortest optical distance between WF and $W'F'$. These lines we have called the "rays." If the wave-surfaces are not spheres, the rays are not in general at right angles to the wave-fronts, and this is an important distinction between crystalline and isotropic media.

If through any point P (Fig. 116), we draw a number of plane wave-fronts, we may, from the results of the last article, construct the positions W_1F_1, W_2F_2, W_3F_3, etc. of these wave-fronts after unit time. Each wave-front must be a tangent plane to the wave-surface drawn with P as centre. Hence the wave-surface is the envelope of all the plane wave-fronts. Its equation may thus be obtained by a purely mathematical process.

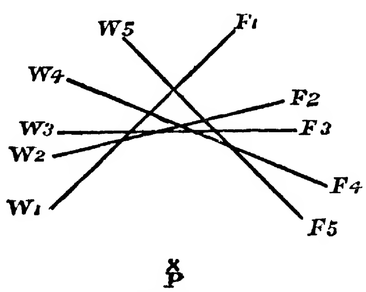


Fig. 116.

The equation to the wave-front is

$$lx + my + nz = v \quad \text{..............................(6)},$$

where v is the distance travelled in unit time, which itself is a function of l, m, n. Any point Q of the wave-surface is a point of intersection of planes, differing from each other in direction by infinitely small quantities. Hence a point x, y, z of the wave-surface must satisfy (6) and also the equation obtained by giving to l, m, n, v small increments dl, dm, dn, dv. Subtracting one of these equations from the other it follows that

$$xdl + ydm + zdn = dv \quad \text{..........................(7)}.$$

There are certain conditions to which the variations of the quantities l, m, n, v, are subject. Thus

$$l^2 + m^2 + n^2 = 1$$

from which we derive

$$ldl + mdm + ndn = 0 \quad \text{..........................(8)},$$

and also as equation (4) continues to hold,

$$\frac{ldl}{v^2 - a^2} + \frac{mdm}{v^2 - b^2} + \frac{ndn}{v^2 - c^2} = Kvdv \quad \text{..............(9)},$$

where

$$K = \frac{l^2}{(v^2 - a^2)^2} + \frac{m^2}{(v^2 - b^2)^2} + \frac{n^2}{(v^2 - c^2)^2}.$$

As there are only two independent parameters of the plane, *e.g.* l and m, it must be possible to express dn and dv in terms of dl and dm. Two equations are sufficient for this purpose, and of the three equations (7), (8), (9) only two can be independent. To express the condition that one of these equations may be obtained as a consequence of the two others, we multiply (8) by A and (9) by B, and add.

Equalizing the factors of dl, dm, dn, and dv, in this combined equation, with the corresponding factors in (7), we find

$$\left.\begin{aligned} x &= Al + B\frac{l}{(v^2-a^2)} \\ y &= Am + B\frac{m}{v^2-b^2} \\ z &= An + B\frac{n}{v^2-c^2} \end{aligned}\right\} \quad\text{..........(10)},$$

$$BKv = 1 \quad\text{..........(11)}.$$

Multiplying (10) by l, m, n, respectively, and adding, we obtain, with the help of (6) and (4),

$$v = A \quad\text{..........(12)}.$$

Squaring and adding the equation (10), and writing r^2 for $x^2+y^2+z^2$, the term containing the product AB disappears in consequence of (4) and we obtain:

$$r^2 = A^2 + B^2K.$$

With the help of (11) and (12) this gives B:

$$B = v(r^2 - v^2) \quad\text{..........(13)}.$$

The first of equations (10) may now be written

$$x = \frac{B + A(v^2-a^2)}{v^2-a^2}\,l = \frac{r^2-a^2}{v^2-a^2}\,vl.$$

$$\left.\begin{aligned} &\text{Hence} && \frac{x}{r^2-a^2} = \frac{vl}{v^2-a^2} = \frac{x-vl}{r^2-v^2}. \\ &\text{Similarly} && \frac{y}{r^2-b^2} = \frac{vm}{v^2-b^2} = \frac{y-vm}{r^2-v^2}, \\ & && \frac{z}{r^2-c^2} = \frac{vn}{v^2-c^2} = \frac{z-vn}{r^2-v^2}. \end{aligned}\right\} \quad\text{..........(14)}.$$

Multiplying these equations by x, y, z, respectively and adding, the quantities l, m, n disappear owing to the relation (6) and we obtain

$$\frac{x^2}{r^2-a^2} + \frac{y^2}{r^2-b^2} + \frac{z^2}{r^2-c^2} = 1 \quad\text{..........(15)}.$$

This is one form of the equation of the wave-surface. Another form is obtained by multiplying both sides of (15) by r^2, and then subtracting $x^2+y^2+z^2$ from the left-hand side, and r^2 from the right-hand side. This leaves

$$\frac{a^2x^2}{r^2-a^2} + \frac{b^2y^2}{r^2-b^2} + \frac{c^2z^2}{r^2-c^2} = 0 \quad\text{..........(16)},$$

a form which is usually more convenient than the former, though analytically identical. Getting rid of the denominators, we find

$$(r^2-b^2)(r^2-c^2)\,a^2x^2 + (r^2-a^2)(r^2-c^2)\,b^2y^2 + (r^2-a^2)(r^2-b^2)\,c^2z^2 = 0.$$

Multiplying out and dividing by the common factor r^2 we find

$$r^2(a^2x^2+b^2y^2+c^2z^2)-a^2(b^2+c^2)x^2-b^2(c^2+a^2)y^2 \\ -c^2(a^2+b^2)z^2+a^2b^2c^2=0 \quad\text{.........(17).}$$

This is an equation of the fourth degree.

91. Ray Velocity. The radius vector of the wave-surface, being the distance through which a disturbance may be considered to have travelled in unit time, measures the velocity along the ray. Calling s the ray velocity, while v is the velocity of plane wave propagation, Fig. 115 shows that if ζ be the inclination of the ray to the wave normal, $v=s\cos\zeta$, or if λ, μ, ν be the direction cosines of the ray,

$$v=s(l\lambda+m\mu+n\nu) \quad\text{.......................(18).}$$

Equation (16) may serve to connect the direction λ, μ, ν with the velocity s. Writing s for r and using $x:y:z=\lambda:\mu:\nu$, we find

$$\frac{a^2\lambda^2}{s^2-a^2}+\frac{b^2\mu^2}{s^2-b^2}+\frac{c^2\nu^2}{s^2-c^2}=0 \quad\text{.................(19).}$$

This equation becomes identical with (4) if we write $1/s$ for v, and $1/a$, $1/b$, $1/c$, for a, b, c respectively and identify l, m, n with λ, μ, ν. This suggests the following construction for obtaining ray velocities similar to that which holds for the wave velocities. Take the ellipsoid

$$\frac{x^2}{a^2}+\frac{y^2}{b^2}+\frac{z^2}{c^2}=1.$$

This ellipsoid has been called the reciprocal ellipsoid. Its semiaxes are equal to the principal velocities. The semiaxes of any plane central section then measure the two possible velocities of rays which are at right angles to the section. The proof of this proposition follows from a reasoning identical with that by means of which (4) has been obtained.

92. Relations between rays and wave normals. The projections of the ray on the three coordinate axes are $s\lambda$, $s\mu$, $s\nu$, and we may substitute these values for x, y, z, in equations (14). We then obtain

$$\left.\begin{aligned}\frac{s\lambda}{s^2-a^2}&=\frac{vl}{v^2-a^2}\\ \frac{s\mu}{s^2-b^2}&=\frac{vm}{v^2-b^2}\\ \frac{s\nu}{s^2-c^2}&=\frac{vn}{v^2-c^2}\end{aligned}\right\} \quad\text{.......................(20)}$$

and $$\lambda:\mu:\nu=\alpha(s^2-a^2):\beta(s^2-b^2):\gamma(s^2-c^2).$$

Squaring and adding, we find:

$$s^2\left[\frac{\lambda^2}{(s^2-a^2)^2}+\frac{\mu^2}{(s^2-b^2)^2}+\frac{\nu^2}{(s^2-c^2)^2}\right]$$
$$=v^2\left[\frac{l^2}{(v^2-a^2)^2}+\frac{m^2}{(v^2-b^2)^2}+\frac{n^2}{(v^2-c^2)^2}\right]\text{.........(21).}$$

We may also by a simple transformation of (20) obtain the three equations

$$vl-s\lambda=s\lambda\frac{v^2-s^2}{s^2-a^2},$$
$$vm-s\mu=s\mu\frac{v^2-s^2}{s^2-b^2},$$
$$vn-s\nu=s\nu\frac{v^2-s^2}{s^2-c^2},$$

and by squaring and adding, find

$$v^2+s^2-2vs\,(l\lambda+m\mu+n\nu)=s^2\,(v^2-s^2)^2\left[\frac{\lambda^2}{(s^2-a^2)^2}+\frac{\mu^2}{(s^2-b^2)^2}+\frac{\nu^2}{(s^2-c^2)^2}\right].$$

Introducing (18) this equation reduces to

$$\frac{1}{s^2-v^2}=s^2\left[\frac{\lambda^2}{(s^2-a^2)^2}+\frac{\mu^2}{(s^2-b^2)^2}+\frac{\nu^2}{(s^2-c^2)^2}\right]\text{.........(22),}$$

and hence from (21)

$$\frac{1}{s^2-v^2}=v^2\left[\frac{l^2}{(v^2-a^2)^2}+\frac{m^2}{(v^2-b^2)^2}+\frac{n^2}{(v^2-c^2)^2}\right]\text{.........(23).}$$

Equation (23) may be used to calculate the ray velocity from the velocity of the corresponding plane wave, while (22) is used when the ray velocity is given and the wave velocity is required.

To determine the angle ζ included between s and v, we may use either

$$\cot^2\zeta=\frac{v^2}{s^2-v^2}\text{ or }\operatorname{cosec}^2\zeta=\frac{s^2}{s^2-v^2}.$$

The former gives, in terms of the quantities defining the plane wave propagation,

$$\cot^2\zeta=v^4\left[\frac{l^2}{(v^2-a^2)^2}+\frac{m^2}{(v^2-b^2)^2}+\frac{n^2}{(v^2-c^2)^2}\right],$$

and the latter in terms of the quantities defining the ray propagation

$$\operatorname{cosec}^2\zeta=s^4\left[\frac{\lambda^2}{(s^2-a^2)^2}+\frac{\mu^2}{(s^2-b^2)^2}+\frac{\nu^2}{(s^2-c^2)^2}\right].$$

The plane containing the direction of vibration OP, Fig. 117, and the wave normal ON contains other vectors which are related to the wave propagation. To express analytically the condition that three vectors should be in the same plane, we make use of the well-known relation between the direction cosines l, m, n, and ξ, η, ζ, of two lines which are at right angles to each other. This condition is

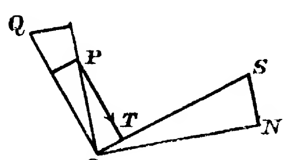


Fig. 117.

$$l\xi+m\eta+n\zeta=0.$$

If ξ, η, ζ, is also at right angles to the vector λ, μ, ν,

$$\lambda\xi + \mu\eta + \nu\zeta = 0,$$

also
$$\xi^2 + \eta^2 + \zeta^2 = 1.$$

Solving these equations, we find

$$\xi = m\nu - n\mu,$$
$$\eta = n\lambda - l\nu,$$
$$\zeta = l\mu - m\lambda.$$

A third vector α, β, γ will be in the same plane with l, m, n and λ, μ, ν, if it is at right angles to ξ, η, ζ. Hence

$$\alpha (m\nu - n\mu) + \beta (n\lambda - l\nu) + \gamma (l\mu - m\lambda) = 0$$

is the required condition. If there is a linear relation between the three vectors, such that

$$\alpha = C\lambda + Fl\,; \;\; \beta = C\mu + Fm\,; \;\; \gamma = C\nu + Fn,$$

the condition is obviously satisfied. Giving to the direction cosines their previous meaning, we have according to (3)

$$\left.\begin{aligned} Dl &= (v^2 - a^2)\,\alpha \\ Dm &= (v^2 - b^2)\,\beta \\ Dn &= (v^2 - c^2)\,\gamma \end{aligned}\right\} \quad \text{..........................(24)},$$

while (23) shows that D is equal to $v\sqrt{s^2 - v^2}$.

Also with the help of (20) and (24)

$$\left.\begin{aligned} E\lambda &= (s^2 - a^2)\,\alpha \\ E\mu &= (s^2 - b^2)\,\beta \\ E\nu &= (s^2 - c^2)\,\gamma \end{aligned}\right\} \quad \text{..........................(25)},$$

where E is written for Ds/v, which is equal to $s\sqrt{s^2 - v^2}$.

Combining (24) and (25) we obtain the following linear relations:

$$E\lambda - Dl = (s^2 - v^2)\,\alpha,$$
$$E\mu - Dm = (s^2 - v^2)\,\beta,$$
$$E\nu - Dn = (s^2 - v^2)\,\gamma,$$

which proves that the three vectors are coplanar. The "ray" therefore lies in the plane which contains the wave normal and the displacement. In Fig. 117 the wave normal and ray are indicated by the direction of the lines ON and OS. We may now prove that the normal PT to the ellipsoid of elasticity at the point at which the direction of the displacement intersects it, lies in the same plane. At the end of the radius vector OP draw the tangent PK to the ellipse of intersection. If OP is a semiaxis, PK is at right angles to OP, and also to PT, the normal to the ellipsoid. Hence PT and PO are in a plane which is at right angles to PK, and hence also at right angles to OP', the second semiaxis of the ellipse. OP' being the normal to the plane containing

PT and PO, every line at right angles to OP' must lie in the same plane. The normal to the section is such a line, hence PO, PT and ON are in the same plane.

The direction cosines of the normal PT are proportional to a^2x, b^2y, c^2z, if x, y, z are the coordinates of P, and hence also proportional to $a^2\alpha$, $b^2\beta$, $c^2\gamma$.

The ray OS is at right angles to that plane section of the reciprocal ellipsoid which has the ray velocity s as one of its semiaxes. If we proceed exactly as in Art. 87 to find the direction cosines of the two semiaxes of the ellipse which is at right angles to λ, μ, ν, we find by equations analogous to (3) that they are proportional to

$$a^2\lambda/(s^2-a^2),\ b^2\mu/(s^2-b^2),\ c^2\nu/(s^2-c^2),$$

and therefore by comparison with (20) proportional to $a^2\alpha$, $b^2\beta$, $c^2\gamma$.

This proves that the semiaxis in question is parallel to PT. Let OQ (Fig. 117) be that semiaxis. The normal to its tangent plane has direction cosines proportional to $\frac{\alpha'}{a^2}$, $\frac{\beta'}{b^2}$, $\frac{\gamma'}{c^2}$, if α', β', γ' fix the direction of OQ. From the ratio $\alpha' : \beta' : \gamma'$ which has just been found, it follows at once that OP is the normal to the tangent plane at Q.

The second semiaxis of the section of the ellipsoid of elasticity passing through OP is coincident in direction with the second semiaxis of the section of the reciprocal ellipsoid, which passes through OQ, because in both cases the semiaxis must be at right angles to the plane of the figure.

93. The direction of displacement. It has been proved that the vectors OP, OS, and ON (Fig. 117) are in the same plane, OP indicating the direction of the displacement. As NS is in the wave-front, the vibration takes place in the direction NS, which is the projection of the ray on the wave-front. The direction of vibration cannot be observed, and the above statement involves therefore something that is theoretical and based on a particular assumption as to the nature of light. That assumption has been introduced by the manner in which the construction of Art. 87 has been carried out, as already explained in the footnote to that article. If we wish to confine our statements to facts capable of experimental verification, we ought to say that the plane of polarization is at right angles to the projection of the ray on the wave-front. The two ways of expressing the facts are identical, if the direction of vibration is at right angles to the plane of polarization.

The planes of polarization may be obtained in a simple manner from the direction of the optic axes. The axes of an ellipse are the bisectors of the angles formed by any two equal diameters, and the planes which are normal to the optic axes intersect the ellipsoid of elasticity in a circle of radius $1/b$. Hence if ON (Fig. 118) represents the normal to the wave-front, and $PP'Q'Q$ the elliptical intersection of the wave-front and the ellipsoid, two lines OP, OP' which are at right angles to the optic axes OH and OH' are of equal length $1/b$. A plane through ON and OH intersects the ellipse in a line OQ at right angles to OP, and a plane through the normal and second optic axis intersects the ellipse along OQ' at right angles to OP'. Hence OQ and OQ' are equal radii of the ellipse. It follows that the planes of polarization are the two bisectors of the planes which pass through the normal and two optic axes respectively.

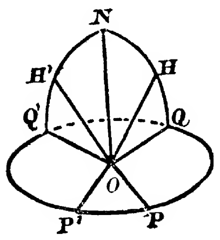


Fig. 118.

The two plane waves propagated in the same direction have their planes of polarization at right angles to each other, but the two planes of polarization belonging to a given *ray* are not at right angles to each other unless the ray coincides with the wave normal. To prove this, we take two directions of vibration α, β, γ, and α_1, β_1, γ_1, which correspond to the same value of λ, μ, ν, the ray velocities being s and s_1, and the corresponding wave velocities v and v_1. According to (25) we have

$$E\lambda = (s^2 - a^2)\,\alpha\,;\; E\mu = (s^2 - b^2)\,\beta\,;\; E\nu = (s^2 - c^2)\,\gamma,$$

$$E_1\lambda = (s_1^2 - a^2)\,\alpha_1\,;\; E_1\mu = (s_1^2 - b^2)\,\beta_1\,;\; E_1\nu = (s_1^2 - c^2)\,\gamma_1,$$

where $$E = s\sqrt{s^2 - v^2} \text{ and } E_1 = s_1\sqrt{s_1^2 - v_1^2}.$$

The cosine of the angle ω between the directions of vibration is

$$\cos\omega = \alpha\alpha_1 + \beta\beta_1 + \gamma\gamma_1,$$

and after substitution, the right-hand side is found to be equal to

$$\frac{EE_1}{s_1^2 - s^2}\left[\left(\frac{\lambda^2}{s^2 - a^2} + \frac{\mu^2}{s^2 - b^2} + \frac{\nu^2}{s^2 - c^2}\right) - \left(\frac{\lambda^2}{s_1^2 - a^2} + \frac{\mu^2}{s_1^2 - b^2} + \frac{\nu^2}{s_1^2 - c^2}\right)\right].$$

With the help of (19) this becomes

$$\frac{EE_1}{s^2 s_1^2} = \frac{\sqrt{(s^2 - v^2)(s_1^2 - v_1^2)}}{ss_1}.$$

Hence if ϕ and ϕ_1 are the angles included between the common direction of the ray and the two wave normals,

$$\cos\omega = \sin\phi\sin\phi_1 \quad\text{.......................(26).}$$

In order that the two directions of vibration should be at right angles to each other, it is therefore necessary that the ray should coincide with one of the wave normals.

94. Shape of the wave-surface. We may now form an idea of the general shape of the wave-surface. If in (17) we put successively $z = 0$, $y = 0$, $x = 0$, we obtain the intersections of the wave-surface with the coordinate planes.

The intersection with the plane of xy is

$$(x^2+y^2)(a^2x^2+b^2y^2)-a^2(b^2+c^2)x^2-b^2(c^2+a^2)y^2+a^2b^2c^2=0,$$

or
$$(x^2+y^2-c^2)(a^2x^2+b^2y^2-a^2b^2)=0.$$

This is separately satisfied by

$$x^2+y^2-c^2=0,$$

and by
$$a^2x^2+b^2y^2-a^2b^2=0.$$

The curve of intersection breaks up therefore into a circle of radius c and ellipses of semiaxes a and b. The circle lies inside the ellipse and does not intersect it, because we have assumed c to be smaller than both a and b.

The intersection with the plane of yz is similarly found to be a circle of radius a and an ellipse of semiaxes b and c. Here the circle lies completely outside the ellipse.

The intersection of the wave-surface with the plane xz splits up into a circle

$$x^2+z^2=b^2,$$

and an ellipse
$$c^2z^2+a^2x^2-a^2c^2=0.$$

The circle and ellipse in this plane intersect (Fig. 119), the four points of intersection being given by

$$\left.\begin{aligned} x &= \pm c\sqrt{\frac{a^2-b^2}{a^2-c^2}} \\ z &= \pm a\sqrt{\frac{b^2-c^2}{a^2-c^2}} \end{aligned}\right\} \text{......(27).}$$

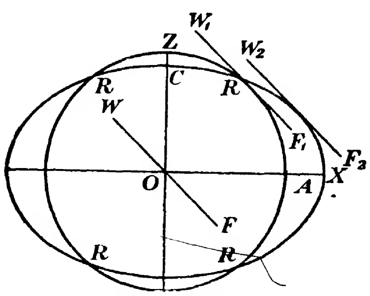


Fig. 119.

Fig. 119 represents in perspective the intersections of the wave-surface with the three coordinate planes, one quadrant only being drawn. The letters attached to the curves of intersection represent the lengths of the semiaxes, thus c, c means a circle of radius c. The complete wave-surface consists of two sheets; an inner sheet, and an outer sheet which meets the inner sheet in four points, one lying in each of the four quadrants of the plane xz. The coordinates of these four points are those given above (27).

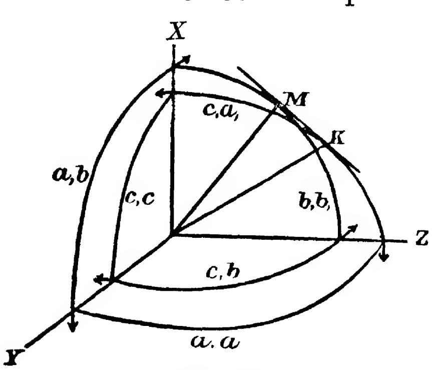


Fig. 120.

The directions of vibration are indicated in the figure by arrows. Each ray which lies in a principal plane coincides with the normal to the wave-surface at one of the points of intersection with it. It therefore coincides with one of the normals of the wave-fronts which may belong to it. Hence according to (26) a ray which lies in one of the principal planes, has its two directions of vibration at right angles to each other. On the elliptical intersections of the wave-front and the principal planes, the direction of vibration, being as proved above the projection of the ray on the tangent plane, must lie in the principal plane. It follows that in the circular intersections of the wave-surface and the principal planes, the direction of vibration is at right angles to the plane.

95. The axes of single ray velocity and of single wave velocity. In general, any straight line drawn in one direction from the origin, intersects the wave-surface in two points, one on each of the two sheets. The intercepts between the origin and the points of intersection measure the optical lengths. There are therefore in each direction in general, two different optical lengths according to the direction of vibration. But the four directions OR (Fig. 119) form an exception, for there is only one point of intersection for each of them, and therefore only one optical length. If a wave is propagated through the crystal, and the ray happens to lie along the direction OR, the difference in optical length between two points along the ray is independent of the direction of the displacements. Adopting the definition of "ray velocity" given in Art. 91 we may call these directions the axes of "single ray velocity." They are not coincident with the optic axes. The latter indicate the directions of single *wave* velocity, the wave being considered to be plane, and normal to one of the optic axes. Remembering that we may obtain the position of a plane wave-front WF, Fig. 119, after unit time, if we construct the wave-surface and draw the tangent planes to that surface parallel to WF, we see from the figure that in general (as we know already), there are two tangent planes W_1F_1, and W_2F_2, which are parallel to each other, and to WF. The lengths of the perpendiculars from O on these planes, measure the wave velocities. If there is a direction in which there is only one wave velocity, the two tangent planes normal to that direction must coincide. There is indeed one tangent plane MK (Fig. 120) in each quadrant, which touches both sheets of the wave-surface simultaneously. Symmetry shows that these tangent planes are parallel to OY, and as they must touch the circle of radius b in the plane of the figure at a point M, OM must be the direction of one of the optic axes. Combining (27) and (5) we find for the cosine of the angle between the optic axes and the axes of single ray velocities,

$$\frac{b^2+ca}{b(a+c)}.$$

If b is equal to either a or c, or if a and c differ but little from each other, the angle is small, and may sometimes be neglected. To form an idea of the error committed in this way, write $c = b(1-\delta)$ and $a = b(1+\epsilon)$, where δ and ϵ are small. We obtain for the cosine of the angle included between the two sets of axes to the second approximation

$$1 - \tfrac{1}{2}\delta\epsilon \qquad \text{.............................(28).}$$

For mica the angle is about 40′. Even in the case of asparagin, a crystal in which the optic axes are nearly 90° apart, the angle between the optic axis and the axis of single ray velocity is less than 2°.

96. Peculiarity of single wave propagation. In general, one ray belongs to each plane wave which is propagated through a crystalline medium, and the radii drawn from O (Fig. 115) to the points of contact of the parallel planes W_1F_1 and W_2F_2 with the wave-surface are the rays belonging to the two waves propagated parallel to WF. When the wave normal coincides with the optic axis, there is only one velocity as we have seen, and inspection of Figure 120 shows that the two rays OM and OK belong to this same wave. But the wave-front WF touches the wave-surface not only at the two points M and K, but along the complete circumference of a circle drawn with MK as diameter. To prove this, we must turn back to equations (14) which determine the points of contact x, y, z, of a plane defined by the direction cosines of its normal (l, m, n) and the wave-surface. To suit our present problem, we must put $m = 0$, $v = b$. The second equation becomes indeterminate and may be satisfied for any value of y, by a proper choice of the two indefinitely small quantities m and $v - b$. The first and third equations are therefore the only ones we need consider. Multiplying the first by l, and the third by n, we obtain

$$\begin{aligned}\frac{xl}{r^2 - a^2} + \frac{zn}{r^2 - c^2} &= v\left(\frac{l^2}{v^2 - a^2} + \frac{n^2}{v^2 - c^2}\right)\\ &= v\left(\frac{l^2}{b^2 - a^2} + \frac{n^2}{b^2 - c^2}\right)\\ &= 0.\end{aligned}$$

The last step follows on substituting for l, n the direction cosines of the optic axis as given in (5). Hence

$$lx(r^2 - c^2) + nz(r^2 - a^2) = 0 \qquad \text{..................(29).}$$

Also $$xl + zn = b \qquad \text{..........................(30),}$$

because the required line of contact must lie in the plane passing through M and touching the circular section. The equations (29) and (30) combined give

$$b(x^2 + y^2 + z^2) = lc^2x + na^2z \qquad \text{..................(31),}$$

which is the equation of a sphere passing through the origin of coordinates.

The point of contact of the tangent plane and the wave-surface is therefore the same as the intersection of the tangent plane and the sphere (31). It follows that the line of contact is a circle. The rays which join O to a point on the circle form a cone, the equation to which may be obtained by multiplying (30) and (31) together, *i.e.*

$$b^2(x^2+y^2+z^2) = (lc^2x + na^2z)(lx + nz).$$

It is a remarkable fact that we may have a plane wave propagation, such that the condition of minimum optical length from a point P is satisfied not only for one direction, but for an indefinite number of directions lying on a cone, CC' (Fig. 121). For any point T which lies either inside or outside the circle forming the base of the cone, the optical length is greater. It should, however, be noticed that if the wave is plane polarized, there is only one ray. The distinction between this case and the general one is therefore that while in general the vibration of an unpolarized wave-front may be decomposed into two, for either of which there is a definite wave velocity and a corresponding ray, the vibrations do not in this special case divide themselves into two components, but to each direction of vibration belongs a different ray, all these rays lying on a cone of the second degree.

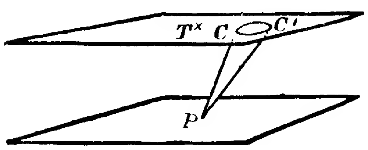


Fig. 121.

97. Peculiarity of a single ray propagation. We may obtain results analogous to the preceding ones, if we try to find the directions of the normals to the tangent planes at the conical point where the direction of single ray velocity cuts the wave-surface. We make use for this purpose of equations (20). With the same notation as before $x = vl$, $z = vn$, are the coordinates of the foot of the perpendicular from the origin to any one of the tangent planes. The ray velocity being constant and equal to b while $\mu = 0$, the second equation is indeterminate and the first and third become

$$\frac{b\lambda}{b^2-a^2} = \frac{x}{r^2-a^2},$$

$$\frac{b\nu}{b^2-c^2} = \frac{z}{r^2-c^2}.$$

With the help of (27) and considering that x and y in that equation are proportional to λ, μ, we find

$$\frac{a^2\lambda x}{r^2-a^2} + \frac{c^2\nu z}{r^2-c^2} = 0,$$

and

$$a^2\lambda x + c^2\nu z = \frac{a^2c^2}{b}.$$

The first equation, with the help of the second, becomes

$$r^2 = b(\lambda x + \nu z).$$

Hence the locus of the foot of the perpendicular from O to the tangent planes at R is the circle formed by the intersection of a sphere with a plane. The equation of the plane shows that it is parallel to OY and touches at R the ellipse ARC, Fig. 119.

98. Wave-surface in uniaxal crystals. The wave-surface in uniaxal crystals takes the shape already indicated by Huygens. If $b = c$ equation (17) reduces to

$$(r^2 - c^2)(a^2x^2 + c^2y^2 + c^2z^2 - a^2c^2) = 0.$$

The surface splits up therefore into the sphere of radius c and the spheroid

$$\frac{x^2}{c^2} + \frac{y^2 + z^2}{a^2} = 1$$

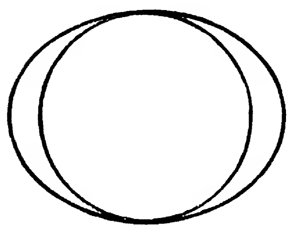


Fig. 122.

Similarly, if $a = b$, the equation of the wave-surface splits up into a sphere of radius a and into the ellipsoid

$$\frac{x^2 + y^2}{c^2} + \frac{z^2}{a^2} = 1.$$

Figs. 122 and 123 represent the two cases, the optic axes being in the first case the axis of X and in the second case the axis of Z. The positions of the axes are determined if we take a, b, c to be always in descending order, but if we drop that supposition, we may take the optic axis of uniaxal crystals to be at our choice either in the direction of X or in the direction of Z. The spheroid is formed by the revolution of the ellipse and circle round the optic axis. The two types of wave surfaces, one having an oblate and the other a prolate spheroid, according as the generating ellipse is made to rotate about its shorter or longer diameter, are illustrated by the case of Iceland Spar and Quartz. The term positive and negative crystals, as applied to crystals similar to Quartz and Iceland spar respectively, is confusing and should be avoided. We may speak instead of crystals which are optically prolate, or optically oblate, and in a discussion relating to optical properties only, where no confusion is possible, we may call them shortly prolate and oblate crystals.

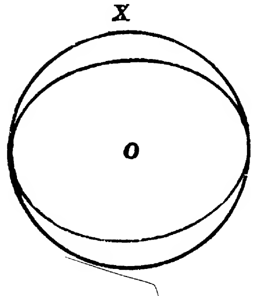


Fig. 123.

99. Refraction at the Surface of Uniaxal Crystals. The refracted waves may, in crystalline media, be constructed exactly as in isotropic bodies, but as the wave-surface consists of two sheets, there are in general two refracted rays. In uniaxal crystals, one sheet of the wave-surface is always a sphere, and hence one of the rays follows the ordinary laws of refraction. This ray is called the ordinary ray, and the ratio of the sines of the two angles of direction is called the ordinary refractive index.

The geometrical construction of the refracted wave-surface and ray may be conducted in a manner similar to that explained in Art. 18, regard being had to the fact that the wave surface now consists of two separate sheets. The plane of the paper representing the plane of incidence, and SO (Fig. 124) the incident wave-normal, draw HB parallel to SO. The perpendicular OH on HB will then lie in the incident wave-front. Construct further with O as centre, the wave-surface, the scale being such that the ratio of HB to the radius OK of the sphere is equal to the ordinary refractive index; this surface will intersect the plane of the paper in a circle, while the curve of intersection of the spheroid is an ellipse. The two refracted wave-fronts are the planes at right angles to the plane of incidence, passing through B and tangent respectively to the sphere and the ellipsoid. The rays are the lines joining O to the points of contact. In the figure BT represents the trace of the refracted extraordinary wave-front, the point of contact with the ellipsoid lying, in this case, outside the plane of incidence.

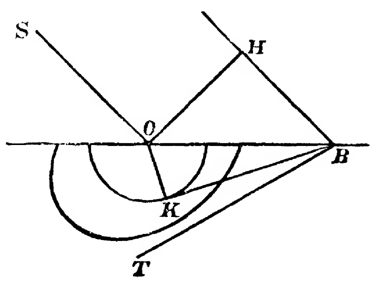


Fig. 124.

It is not necessary to obtain the general equation, giving the direction of the refracted ray, and we may treat a few special cases separately.

(*a*) *The optic axis of the crystal at right angles to the plane of incidence.* The trace of the wave-surface on the plane of incidence is in this case a circle, and the refracted ray may by symmetry be seen to lie in the plane of incidence. Hence the rays follow the ordinary law of refraction. In oblate crystals the outer circle of radius a belongs to the extraordinary ray, and its angle of refraction is greater. The reverse holds for prolate crystals. For oblate crystals, the ratio of the sines for the extraordinary ray is with the unit time chosen $1/a$. Calling this μ_e we may write for the equation to the wave-surface

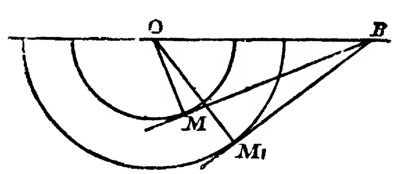


Fig. 125.

$$\mu_0^2 x^2 + \mu_e^2 (y^2 + z^2) = 1,$$

where μ_0 is the refractive index of the ordinary ray. The extraordinary refractive index μ_e has obtained its name and significance from the optic behaviour of the extraordinary ray in the general case we are now considering. In the case of prolate crystals, the equation of the wave-surface in terms of the principal refractive indices becomes

$$\mu_e^2(x^2+y^2)+\mu_0^2 z^2 = 1,$$

the axis of z being now the optic axis.

(*b*) *The optic axis is in the surface and plane of incidence.* The refracted rays are both in the plane of incidence. From the projective properties of the ellipse,

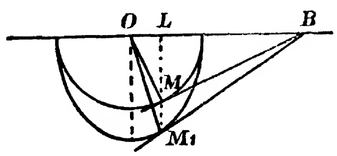


Fig. 126.

$$\frac{LM}{LM_1}=\frac{c}{a},$$

and if the angles of refraction of the *wave normals* are ϕ and ϕ_1,

$$\frac{\tan\phi}{\tan\phi_1}=\frac{\tan OBM}{\tan OBM_1}=\frac{LM}{LM_1}=\frac{\mu_e}{\mu_0},$$

an equation which holds for both prolate and oblate crystals.

If the angles of refraction of the *rays*, OML and OM_1L, be denoted by r and r_1, we obtain similarly

$$\frac{\tan r}{\tan r_1}=\frac{\mu_0}{\mu_e}.$$

(*c*) *The optic axis is perpendicular to the refracting surface* If P_1AQ_1 be the trace of the ellipsoid on the plane of incidence, and if we construct a circle with $P_1Q_1=2a$ as diameter, we have, writing ϕ_1 for the angle LBM, which is the angle of refraction of the extraordinary wave,

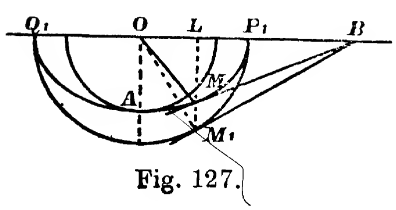


Fig. 127.

$$\frac{\tan\phi_1}{\tan LBM_1}=\frac{LM}{LM_1}=\frac{c}{a}.$$

But $$\sin LBM_1=\frac{OM_1}{OB},$$

and if, for OM_1, we put its value a, and for OB, $1/\sin i$,

$$\tan\phi_1=\frac{c\sin i}{\sqrt{1-a^2\sin^2 i}}:$$

or introducing $$\mu_0=1/c;\ \mu_e=1/a,$$

$$\tan\phi_1=\frac{\mu_e\sin i}{\mu_0\sqrt{\mu_e^2-\sin^2 i}}.$$

And similarly if r_1 is the angle of refraction of the extraordinary ray,

$$\tan r_1=\frac{\mu_0\sin i}{\mu_0\sqrt{\mu_e^2-\sin^2 i}}.$$

(*d*) *The incident wave-front is parallel to the surface.* Let the plane of the paper (Fig. 128) contain the optic axis. The refracted extraordinary ray lies along OM, where M is the point of contact of the spheroidal portion of the wave-surface with a plane drawn parallel to the surface. The ordinary ray coincides, of course, with the normal ON. To determine the angle between the two rays, which is also the angle of refraction of the extraordinary ray, we must obtain an expression for the angle between the radius vector OM of an ellipse, and the normal to its tangent at M. If θ be the angle between the optic axis and the surface which is equal to the angle between ON and OH, the major axis of the ellipse, and γ be the angle between OM and OH, we have by the properties of the ellipse,

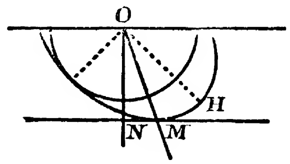


Fig. 128.

$$\tan\theta = \frac{a^2}{c^2}\tan\gamma.$$

Hence if r is the angle of refraction of the extraordinary ray,

$$\tan r = \tan(\theta - \gamma) = \frac{\tan\theta - \tan\gamma}{1 + \tan\theta\tan\gamma}$$

$$= \frac{(a^2 - c^2)\tan\theta}{a^2 + c^2\tan^2\theta}$$

$$= \frac{\mu_0^2 - \mu_e^2}{\mu_0^2\cot\theta + \mu_e^2\tan\theta}.$$

100. Direction of vibration in uniaxal crystals. The rule that the direction of vibration is in the direction of the projection of the ray on the wave-front shows at once that on the spheroidal portion of the wave-front, the direction of vibration must be in a plane containing the optic axis. As the condition (Art. 93) under which the two vibrations along the same ray are at right angles to each other always holds in uniaxal crystals, we may say that the ordinary ray is always polarized in a principal plane, and the extraordinary ray at right angles to that plane.

101. Refraction through a crystal of Iceland Spar. A crystal of Iceland Spar is a rhomb (Fig. 129). The parallelograms forming its six faces have sides which include angles of 102° and 78° respectively. The faces are inclined to each other at angles of 105° and 75°. There are two opposite corners A and B at which the three edges all form obtuse angles at 102° with each other. The optic axis is parallel to the line drawn through one of these corners A, and equally inclined to the three faces.

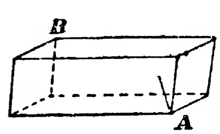


Fig. 129.

Double refraction may easily be exhibited by placing such a rhomb on a white sheet of paper on which a sharp mark is drawn. When this mark is looked at from above through the crystal, it appears double, and if the crystal be turned round, one image seems to revolve round the other. Let O, Fig. 130, be the mark, the images of which are observed. To trace the image formed by the extraordinary rays, construct a wave-surface to such a scale that the spheroid touches the upper surface of the crystal. If T is the point of contact, a ray OT is refracted outwards along the normal TM, because at T the tangent plane to the wave-surface and the surface are coincident. The refraction is therefore the same at that point as for a wave incident normally.

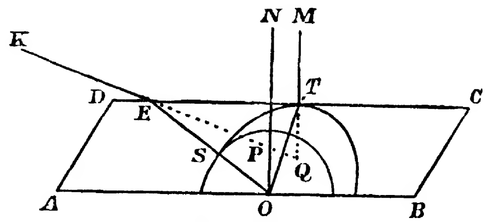


Fig. 130.

A ray OS parallel to the optic axis intersects the face at a point E, and is refracted along some direction EK Disregarding aberrations, the intersection Q of KE and TM gives the extraordinary image. As there can be no distinction between an ordinary and an extraordinary ray along the optic axis, the ordinary image P is obtained by the intersection of the same line EK with the normal ON, on which the ordinary image must lie. The figure shows that this ordinary image lies nearer to the surface than the extraordinary one, and if the crystal be turned round the point O, the image Q travels in a circle round P. The vertical plane containing P and Q contains also the optic axis, and the ordinary image is therefore polarized in the plane which passes through the two images, the extraordinary image being polarized at right angles to it.

102. Nicol's Prism. A Nicol's prism, or, as it ought to be more appropriately called, a Nicol's rhomb, is one of the most useful appliances we have for the study of polarization. Let Fig 131 represent the section of a long rhomb of Iceland Spar, passing through the optic axis, and LL' an oblique section through it. If the rhomb be cut along this section and then recemented together by means of a thin layer of Canada balsam, only rays polarized at right angles to the principal plane are transmitted through it, if the inclination of the section LL' has been properly chosen. An unpolarized ray is refracted at the surface, and separated into two, the extraordinary ray being bent less away from the original direction. The ordinary ray falls therefore more obliquely on the surface of

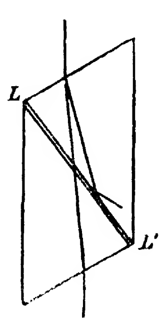


Fig. 131.

separation LL'. The velocity of light in Canada balsam being intermediate between that of the two sets of waves in Iceland Spar, the inclination of LL' may be adjusted so that the ordinary ray is totally reflected, while the extraordinary ray passes through the combination. Fig. 132 shows in perspective how the plane of division is cut through the rhomb. When the end face $ABCD$ of the rhomb is a parallelogram and parallel to one of the cleavage planes, the inclination of the section must be such that the side BB' of the rhomb is about 3·7 times as long as one of the sides of the end faces. It is difficult to secure crystals of Iceland Spar which are sufficiently long to give, under these conditions, a beam of such cross section as is generally required in optical work. The angular space through which the Nicol prism is effective in polarizing light is determined by the fact that if the incidence on the face LL' is too oblique, the extraordinary ray is totally reflected as well as the ordinary ray, and if not oblique enough, the ordinary ray can pass through. The field of view containing the angular space thus limited when the prism is cut according to the above directions, is about 30°. If it is not necessary to have so wide a field of view, shorter lengths of crystals can be used by cutting the end face $ABCD$, so as to be more nearly perpendicular to the length. Sometimes that face is even inclined the other way. A field of view of 25° may thus be secured with a ratio of length to breadth of 2 to 5. Artificial faces at the end have, however, the disadvantage of deteriorating more quickly than cleavage planes.

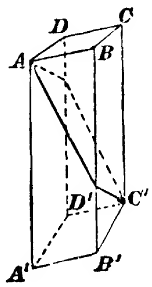


Fig. 132.

Foucault constructed a rhomb in which a small thickness of air is introduced in place of the Canada balsam. The prism need then be barely longer than broad, but the field of view is reduced to 7°.

103. Double Image Prisms. It is sometimes convenient to have two images of a source near together, achromatic as far as possible, and polarized perpendicularly to each other. An ordinary prism made of Iceland spar or quartz cannot be used on account of the colour dispersion, but if a prism of quartz be achromatised by means of a prism of another material, the desired result may be obtained. If glass is chosen for the material of the second prism, the achromatism is only complete for one of the images, but for many purposes it is sufficiently perfect for the second image also. The purpose is better obtained by prisms, like that of Rochon, in which the same material is used for both prisms, but turned differently with respect to the optic axis. In Rochon's

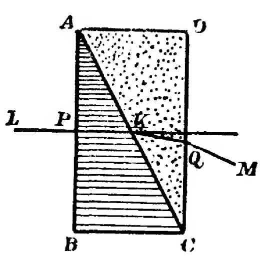


Fig. 133.

arrangement the optic axis of the first prism ABC (Fig. 133) is parallel to the normal BC, this being indicated in the figure by the direction of the shading. A ray LP incident normally is propagated without change of direction. The axis of the second prism ACD is at right angles to the plane of the figure, and double refraction takes place at K, one ray being propagated in the normal direction as before, but the extraordinary ray being refracted along KQ and, on passing out of the prism, along QM. The achromatism is complete for the image formed by the ordinary ray, and nearly complete for the other. In the prism of Wollaston (Fig. 134), the axis of the first prism is parallel to AB and that of the second at right angles to the plane of the figure; the path of the rays is indicated in the figure.

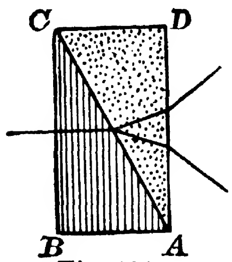


Fig. 134.

104. Principal Refractive Indices in biaxal crystals. If refraction takes place at the surface of a biaxal crystal, and the plane of incidence is one of the principal planes (*e.g.* the plane of YZ), both rays lie in the plane of incidence. A plane wave-front incident at O must, after refraction, touch a circle of radius a, and an ellipse of semiaxes b and c which form the intersection of the wave-surface with the plane of YZ. One of the rays follows the ordinary law of refraction, while the angle of refraction of the other ray may be obtained as in case (c), Art. 99. The refractive index of the rays belonging to the circular section is $1/a$; similarly for planes of incidence coincident with the planes of XZ and of XY, we should have always one ray following the ordinary law, the corresponding refractive indices being $1/b$ and $1/c$. These three quantities are therefore called the principal refractive indices, and all quantities relating to the wave-surface may be expressed in terms of them. Thus for the direction cosines of the optic axes, we have from (5), if the refractive indices be denoted by μ_1, μ_2, μ_3,

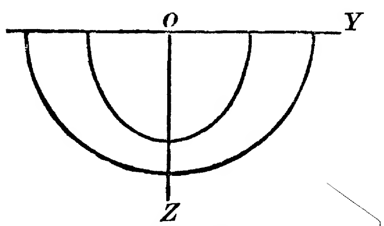


Fig 135.

$$l_1 = \pm \frac{\mu_3}{\mu_2}\sqrt{\frac{\mu_2^{\,2}-\mu_1^{\,2}}{\mu_3^{\,2}-\mu_1^{\,2}}}\,;\ m_1 = 0\,,\ n_1 = \pm \frac{\mu_1}{\mu_2}\sqrt{\frac{\mu_3^{\,2}-\mu_2^{\,2}}{\mu_3^{\,2}-\mu_1^{\,2}}},$$

and for the direction cosines of the rays of single ray velocity, using (27),

$$\lambda = \pm\sqrt{\frac{\mu_2^{\,2}-\mu_3^{\,2}}{\mu_3^{\,2}-\mu_1^{\,2}}}\,;\ \mu = 0\,;\ \nu = \pm\sqrt{\frac{\mu_3^{\,2}-\mu_2^{\,2}}{\mu_3^{\,2}-\mu_1^{\,2}}}.$$

105. Conical Refraction. Two cases of refraction in biaxal crystals have a special interest. If a wave-front WF is incident on a plate cut out of the crystal at an angle such that the refracted wave-front $HKLM$ is normal to an optic axis, the ray PD may, according to the direction of vibration, be refracted along any direction lying on the surface of the cone investigated in Art. 96, the cone intersecting the wave-front inside the crystal in a circle (Fig. 136).

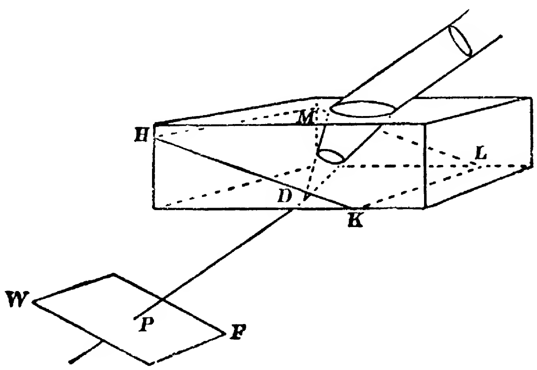


Fig. 136.

If the wave-front WF contains a number of coincident rays, having their planes of polarization symmetrically distributed in all directions, the refracted rays form the surface of a cone of the second degree which becomes a cylinder on emergence at the upper surface. This interesting result was first deduced theoretically by Sir Wm. Hamilton, from the shape of the wave-surface, and was afterwards experimentally verified by Lloyd. To illustrate it experimentally, we may take a plate (Fig. 137), cut so that its face is equally inclined to both axes. An opaque plate PQ with a small aperture O, covers the side on which the light is incident. A second plate $P'Q'$ transmits light through a small hole at O', which, if properly illuminated, may be considered to act as a source of light. If now PQ be moved along the face of the crystal, a direction $O'O$ may be found such that if the original light is unpolarized, the ray $O'O$ splits into a conical pencil, which may be observed after emergence at AB. This phenomenon is called "internal conical refraction" to distinguish it from another similar effect which takes place when a ray travels along an axis of single ray velocity.

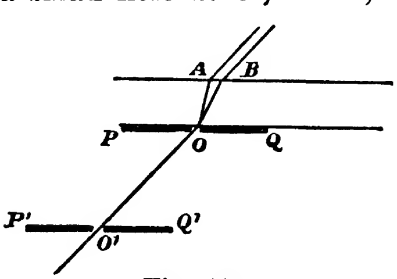


Fig. 137.

We may always follow the refraction of a ray belonging to a certain wave-surface and incident internally on the face of a crystal by considering it to be part of a parallel beam. The wave-front belonging to this parallel beam would be the plane which touches the incident wave-surface at the point of incidence. If now a ray HR (Fig. 138) travels inside a crystal along the axis of single ray velocity, there is an infinite number of tangent planes to the wave-surface at the point R, the normals of the tangent planes forming a cone HKL with a circular section at right angles to HR.

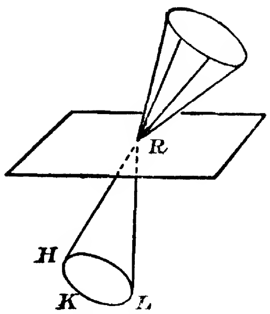


Fig. 138.

To each of these normals corresponds a separate ray on emergence and each ray has its own plane of polarization. The complete cone can only be obtained on emergence if all directions of vibration are represented in the incident ray.

Fig. 139 shows how the phenomenon of external conical refraction may be illustrated experimentally. A plate of arragonite has its surfaces covered by opaque plates, each having an aperture. If one of these plates be fixed and the other is movable, a position may be found of the apertures O and O' such that only such light can traverse the plate as passes along the axis of single ray velocity. The rays on emergence are found to be spread out and to form the generating lines of a cone. But as any ray after passing through a plate must necessarily be parallel to its original direction, it follows that to obtain the emergent cone, the incident beam must also be conical. This may be secured by means of a lens LL' arranged as in the figure. Those parts of the incident beam forming a solid cone which are not required, do not travel inside the crystal along OO' and hence are cut off by the plate covering the upper surface.

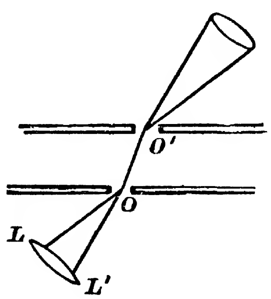


Fig 139.

106. Fresnel's investigation of double refraction. Fresnel's method of treating double refraction which led him to the discovery of the laws of wave propagation in crystalline media, though not free from objection, is instructive, and may be modified so as to bring it into harmony with our present views on the cause of optical dispersion in a material medium (see Chapter XI).

Consider a particle P attracted to a centre O with a force a^2x when the particle lies along OX, and a force b^2y when it lies along OY. The time of oscillation, if the particle has unit mass, is, by Art. 2, $2\pi/a$ or $2\pi/b$ according as the oscillation takes place along the axis of X or along the axis of Y. When the displacement has components both along OX and along

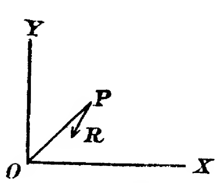


Fig. 140.

OY, the components of the force are a^2x and b^2y, and the resultant force is
$$R = \sqrt{a^4x^2 + b^4y^2}.$$

The cosines of the angles which the resultant makes with the coordinate axes are a^2x/R and b^2y/R. The direction of the resultant force is not the same as that of the displacement, the direction cosines of which are x/r and y/r. The cosine of the angle included between the radius vector and the force is found in the usual way to be
$$\frac{a^2x^2 + b^2y^2}{Rr},$$
and the component of the force along the radius vector is
$$(a^2x^2 + b^2y^2)/r.$$

If we draw an ellipse $a^2x^2 + b^2y^2 = k^2$, where k is a constant having the dimensions of a velocity, the normal to this ellipse at a point P, having coordinates x and y, forms angles with the axes, the cosines of which are in the ratio a^2x to b^2y, hence the force in the above problem acts in the direction ON of the line drawn from O at right angles to the tangent at P. The component of the force along the radius vector is k^2/r, and the force per unit distance is k^2/r^2, so that if the particle were constrained to move on the radius vector OP, its period would be $2\pi r/k$. The ratio r/k depending only on the direction of OP, our result is independent of the particular value we attach to k.

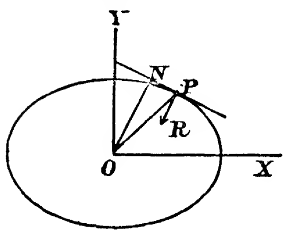


Fig. 141.

If we extend the investigation to three dimensions, the component of attraction along OZ being c^2z, we obtain the same result, and the component of force acting along any radius vector OP per unit length is k^2/r^2, where r is the radius drawn in the direction of OP to the ellipsoid
$$a^2x^2 + b^2y^2 + c^2z^2 = k^2.$$

If the displacement is in any diametral plane HPK of this ellipsoid (Fig. 142), the normal PN does not in general lie in this plane, and the projection of PN on the plane does not pass through O, unless OP is a semiaxis of the ellipse HPK. In the latter case, PL the tangent to the ellipse in the diametral plane, is at right angles to PO and to PN, and hence the plane containing PO and PN is normal to the plane of the section.

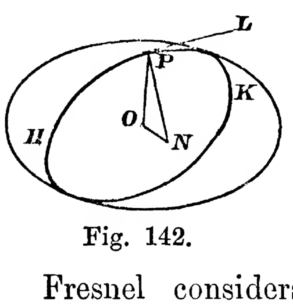


Fig. 142.

Fresnel considers the condition under which a plane wave propagation is possible in a crystalline medium. The investigation in Art. 12 has shown that the accelerations of any point in a plane distortional wave of homogeneous type, are the same as those due to central attracting forces. It is also clear that a plane polarized wave cannot be transmitted as a single wave unless the force of restitution is in the direction of the displacement. If we disregard longitudinal waves as having no reference to the phenomena of light, we need only consider that component of the force which acts in the plane of the wave. This consideration leads to Fresnel's construction. For if we take the ellipsoid

$$a^2x^2 + b^2y^2 + c^2z^2 = 1,$$

which, as we now see, is quite appropriately called the ellipsoid of elasticity, a central section parallel to the wave-front gives an ellipse which, by its principal axes, indicates the two directions of displacement which are compatible with a transmission of a single plane wave. The periods of oscillation are proportional to the axes of this section, and as for a given wave-length the periods of oscillation are inversely proportional to the velocity of transmission, it follows that the velocities of the plane waves parallel to the section are inversely proportional to the axes of the ellipse of intersection. We have thus arrived at the construction which has formed the starting point of our discussion of the phenomena of double refraction (Art. 84).

The direction of the elastic force for any displacement being parallel to the normal to the ellipsoid of elasticity, drawn at the point at which the direction of the displacement intersects the ellipsoid, the proposition proved in Art. 89 shows that the four vectors representing the direction of vibration, the elastic force, the ray and the wave-normal are coplanar.

CHAPTER IX.

INTERFERENCE OF POLARIZED LIGHT.

107. Preliminary Discussion. If a plane unpolarized wave enters a plate of a doubly refracting substance, the two waves inside the crystal travel with different velocities and in slightly different directions, but on emergence both waves are refracted so as again to become parallel to their original directions. If the wave was originally polarized and the plane of polarization be gradually turned, it is found that there are two positions at right angles to each other for which there is only one emergent wave, but in general two overlapping waves polarized perpendicularly will leave the crystal. Their combination forms a beam of light which is polarized elliptically, but for analytical purposes it is often more convenient to treat the two waves temporarily as having a separate existence. If the emergent system be passed through a Nicol prism or any polarizing arrangement, which transmits oscillations in one direction only, one component of the two displacements will be eliminated, and the ultimate wave is polarized in a definite direction determined by the position of the Nicol. The two waves which are now combined have suffered different retardations in their passage through the crystalline plate, and interference is in consequence possible. The description and explanation of the resulting colour effects form the subject of this chapter.

In general, a wave of polarized light incident on a doubly refracting plate becomes polarized elliptically. The axes of the ellipses vary with the wave-length and the thickness of material travelled through, hence also with the direction of the incident light, and the ellipse may, in particular cases, become a straight line or a circle. If the emergent light is examined through a Nicol prism or any arrangement which transmits oscillations in one direction only, colour effects are observed.

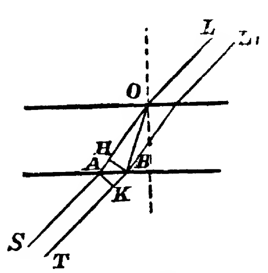


Fig. 143.

It is clear that any interference effect must depend on the difference of phase in the two overlapping emergent waves. Let LO (Fig. 143) be an incident ray, forming part of a parallel beam, OA and OB the refracted *wave normals*, AS and BT the emergent wave normals. Draw AK at right angles to BT and BH at right angles to OA. Imagine a second incident ray, parallel to the first, and at such a distance that the wave normal which is parallel to OA passes through B, and is refracted outwards along BT; then from the principle of wave transmission it follows that the optical length of BK is the same as that of AH. In the emergent wave-front, the difference in optical length is therefore $\frac{OB}{v_1} - \frac{OH}{v_2}$, where v_1 and v_2 are the velocities of the waves along OB and OH respectively. (The unit time is still taken to be such that the velocity of light in vacuo is one.) The angle between OB and OA is small, and if we neglect its square, we may write $OB = OH$. The difference in optical length is therefore $\rho\left(\frac{1}{v_1} - \frac{1}{v_2}\right)$, where ρ is the length of that wave normal inside the plate, which lies nearest to the plate normal. Unless the incidence is very oblique, it makes no difference, to the degree of approximation aimed at, along which wave normal ρ is measured, but for the sake of definiteness, we adhere to the specified meaning of ρ. If OB and OA represent the refracted *rays*, we argue similarly that by Fermat's principle, optical lengths may be measured along a path near the real one, committing only an error of the second order. The optical length for the ray of velocity s_2 might therefore be measured either along its real path OA or along its neighbour $OB + BK$, ending, of course, in the same wave-front. We may therefore also express the difference in optical length as $t\left(\frac{1}{s_1} - \frac{1}{s_2}\right)$, where t is the length of ray inside the crystal and s_1, s_2, are the ray velocities. We may, according to convenience, use either one or the other two forms, which are both approximate only. Which of these is the more accurate in a particular case depends on the question as to whether the angle between the two ray velocities or between the two wave normals is the smaller. In the neighbourhood of the optic axes, it is preferable to refer the relative retardation to the wave normals.

108. Intensity of illumination in transmitted light. Consider polarized light with its direction of vibration along OP (Fig. 144), falling normally on the surface of a crystal which divides the wave into two portions, one vibrating along OX and one along OY. After traversing the thickness of the plate, the two waves emerge normally with a difference of phase δ depending on the difference in optical length of the two wave normals inside the crystal. If the amplitude of the incident light is one, the emergent waves have amplitudes $\cos\alpha$, $\sin\alpha$, if α is the angle between OP and OX, there being a difference in phase δ between them. If now the emergent beam be examined through a Nicol prism called the "analyser," transmitting light only which vibrates along OA, the

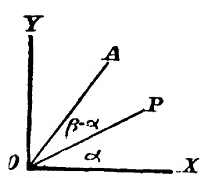


Fig. 144.

component k_1 of the transmitted light due to that portion which in the crystal had OX for its direction of vibration, is $k_1 = \cos\beta\cos\alpha$; similarly $k_2 = \sin\beta\sin\alpha$ is that component of the light which, having OY for the direction of vibration inside the crystal, is capable of traversing the analyser.

Two rays of amplitude k_1 and k_2 and phase difference δ, polarized in the same direction, have a resultant, the intensity of which is

$$k_1^2 + k_2^2 + 2k_1k_2\cos\delta,$$

for which we may write

$$(k_1 + k_2)^2 - 4k_1k_2\sin^2\frac{\delta}{2}.$$

Substituting the values of k_1 and k_2 the intensity of the emergent beam becomes

$$I = \cos^2(\beta - \alpha) - \sin 2\alpha \sin 2\beta \sin^2\frac{\delta}{2} \quad \text{...............(1).}$$

All colour or interference effects shown by crystalline plates when examined by polarized light, depend on the application of this formula. So long as there is only one parallel beam, the plate having the same thickness everywhere, all the quantities are constant, and the plate appears uniformly illuminated. Important particular cases are those in which the Nicols are either parallel ($\alpha = \beta$), or crossed at right angles $\left(\beta - \alpha = \pm\frac{\pi}{2}\right)$.

In the first case we have

$$I_0 = \left(1 - \sin^2 2\alpha \sin^2\frac{\delta}{2}\right),$$

and in the second $\qquad I_1 = \sin^2 2\alpha \sin^2\frac{\delta}{2},$

which shows that $\qquad I_0 + I_1 = 1.$

This relation is a particular case of the general law that if for any value of α and β, $I = I_A$, and if I becomes I_B on turning either the analyser or polarizer through a right angle, then $I_A + I_B = 1$.

We may convince ourselves that this is true without having recourse to the equations. The light falling on the analysing Nicol is partly transmitted and partly deviated to one side, the two portions making up together the incident light which is supposed to be white. On rotating the Nicol through a right angle the transmitted and deviated portions are interchanged so that the complementary effect must be observed.

When white light passes through the plate, the relative proportion of different colours is not in general preserved because δ depends on the wave-length. If a is the amplitude of light of a particular wave-length, so that white light may be represented by Σa^2, the light transmitted through the system is

$$\cos^2(\alpha-\beta)\,\Sigma a^2 - \sin 2\alpha \sin 2\beta\,\Sigma\left(a^2 \sin^2\frac{\delta}{2}\right).$$

The first term represents white light of intensity proportional to $\cos^2(\alpha-\beta)$, and the second term represents coloured light. The relative proportion of the different wave-lengths is not affected by a change in α or β, but the total colour effect may change because the product $\sin 2\alpha \sin 2\beta$ may be either positive or negative. In the first case, we get a certain colour, in the second, white light minus that colour, *i.e.* the complementary colour. We distinguish two special cases.

Case 1. The Nicols are crossed so that $\alpha-\beta=\frac{\pi}{2}$. Here we have

$$I=\sin^2 2\alpha\,\Sigma\left(a^2 \sin^2\frac{\delta}{2}\right).$$

The colours are most saturated in this case, because there is no admixture of white light. As the axes of x and y are fixed in the crystal, we may vary α without change of $\alpha-\beta$ by turning the crystalline plate in its own plane. There will then be four places of maximum intensity at which $\alpha=45°$ or an odd multiple thereof, and four places of zero intensity at which α is a multiple of 90°.

Case 2. The Nicols are parallel so that $\alpha=\beta$. Here we have

$$I=\Sigma a^2 - \sin^2 2\alpha\,\Sigma\left(a^2 \sin^2\frac{\delta}{2}\right).$$

The colour here is always complementary to that in the previous case for the same value of α, the light being white when α is a multiple of a right angle, and most saturated when α is an odd multiple of 45°.

If for any value of α and β, the crystal is turned in its own plane, there are eight positions at which $\sin 2\alpha \sin 2\beta$ vanishes; these occur whenever one of the axes OX and OY coincides with the principal planes of either the polarizing or analysing Nicol. In these positions of the crystal, the light is white, and on passing through these positions, the colour changes into its complementary.

109. Observations of colour effects with parallel light. The general experimental arrangement by means of which the colour effects of polarized parallel light may be shown, is sketched diagrammatically in Fig. 145 MM' is a mirror reflecting the light from the sky, N_1 and N_2 the polarizing and analysing Nicols, CC' is the crystalline plate.

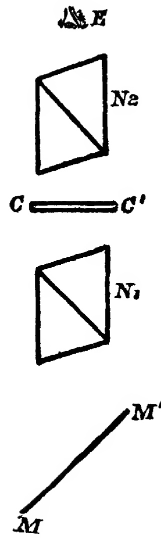


Fig. 145.

The field of view of a Nicol prism is much restricted by the increased distance of the eye from the polarizer N_1. Hence when light from a distant source, such as the sky, passes through both Nicols, only such waves reach the eye as subtend a small angle. The eye at E, focussed for infinity, receives light therefore which has passed through the crystal nearly in the normal direction, and the crystal appears coloured with a uniform tint. If the eye is focussed on the crystal, the colours are not so pure because the different rays leaving the same point of the crystal have traversed it at different inclinations, but when the crystal is thin, so that the relative retardation is only a few wave-lengths, a small variation in direction does not produce much effect on the colour, and therefore the colours are seen with the eye focussed on the plate, nearly as well as with the eye adjusted for parallel light.

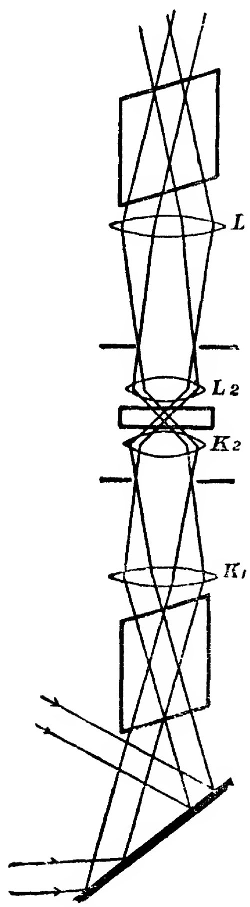


Fig. 146.

An interesting variation of the experiment may be made if the analysing Nicol is replaced by a double image prism; two partially overlapping images of the plate are then seen. The images are coloured where they are separate, but white where they overlap, showing that the colours are complementary.

110. Observations with light incident at different angles. If the field of view is enlarged so as to include rays which have traversed the crystal at sensibly different angles, the effects are more complicated because they depend on the part of the crystal looked at, so that the plate appears to be covered with a pattern of coloured bands. To realize experimentally the necessary increase of the field of view, we may look at the crystal plate through an inverted telescopic system consisting of two lenses L_1 and L_2, placed so as to *diminish* angular distances. The different parallel pencils which have passed through the crystal, pass out of this system with their axes more nearly parallel, so that they may now be sent through a Nicol. A similar telescopic system $K_2 K_1$ serves to *increase* the angular deviation of the rays which have passed through the polarizing Nicol. The thickness of the plate used ought now to be rather larger because it is desired to bring out the differences which are due to variations of length of paths and inclination. When crystals are examined in this fashion, it is generally said that convergent or divergent light is used, but it must be clearly understood that the rays of light which are brought together on the retina traversed the crystal as a parallel pencil. So long as the eye is focussed for infinity, the sole distinction between this case and the previous one, lies in the increase of the field of view.

111. Uniaxal Plate cut perpendicularly to the axis. In order to show how the equation (1) is to be applied to the explanation of the interference pattern under the experimental conditions of the last article, we may treat first the simple case of a plate cut normally to the axis of a uniaxal crystal. An eye E looking in an oblique direction through such a plate (Fig. 147) receives rays which have passed through lengths of path in the crystal, which only depend on the angle between the line of vision and the normal to the plate. Hence the retardation δ is the same along a circle drawn on the surface of the crystal, having its centre coincident with the foot of the perpendicular from the eye to the plate. As the colour effects depend on δ, the field of view is traversed by coloured circular rings. A line along which δ is constant is called an isochromatic line, but the term *isochromatic* here includes the complementary colour. The illumination is not constant along an isochromatic line on account of the variations of α and β. In Fig. 148 $ABCD$ represents the plate, N the foot of the perpendicular from the eye to the plate. If the line of vision passes through the point O, NO is the trace of the plane of incidence, and this plane also contains the optic axis. The two directions of vibration of the ray inside the crystal are therefore NO and the line at right angles to it, and to make equation (1) apply, we must put the axes of X and Y along those directions. The circle drawn through O with N as centre is an isochromatic line. The polarizing and analysing directions remain fixed in space, while the coordinate axes revolve with the point O round N. Whenever either $\sin 2\alpha = 0$ or $\sin 2\beta = 0$, the colour term disappears and we obtain therefore in general four diameters along which there is no coloration. The lines drawn along these directions are called achromatic lines.

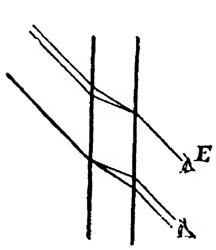


Fig. 147.

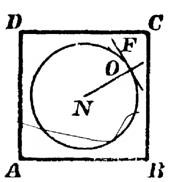


Fig. 148.

We consider three cases.

Case 1. The Nicols are crossed, *i.e.* $\alpha - \beta = \dfrac{\pi}{2}$.

The intensity as before is given by

$$I = \sin^2 2\alpha \Sigma \left(a^2 \sin^2 \frac{\delta}{2}\right).$$

There are two lines at right angles to each other, along which the intensity is zero, these lines coinciding with the directions of the planes of polarization of the analysing and polarizing Nicols. The intensity is greatest at an angle of 45° from these lines. The field is traversed by rings of varying colours, or in the case of homogeneous light, by coloured rings of varying intensity, the dark rings corresponding to the positions at which the phase retardation δ is a multiple of four right angles. The whole appearance consists therefore of a number of concentric rings with a dark cross, as shown in the photograph reproduced in Plate II., Fig. 1. The cross widens out away from the centre and each of its branches is sometimes referred to as a "brush."

Case 2. The Nicols are parallel, *i.e.* $\alpha = \beta$.

The intensity is

$$\Sigma a^2 - \sin^2 2\alpha \Sigma \left(a^2 \sin^2 \frac{\delta}{2}\right),$$

and the whole effect is complementary to that observed in the first case. The rings are now crossed by *bright* brushes. Plate II., Fig. 2 shows the appearance.

Case 3. This includes all positions of the analyser and polarizer in which these are neither parallel nor crossed. There are four achromatic lines corresponding to $\alpha = 0$ and $\alpha = \frac{\pi}{2}$; $\beta = 0$ and $\beta = \frac{\pi}{2}$.

Along an isochromatic circle, the colour changes into its complementary (or for homogeneous light, a minimum of light changes into a maximum) on crossing one of the achromatic lines. This is shown in Plate II., Fig. 3 which is also a reproduction of a photograph. When either the axis of x or the axis of y falls within the acute angle formed by the directions of the analyser and polarizer, the product $\sin 2\alpha \sin 2\beta$ is negative so that the maxima of light are brighter and the minima less dark. The field is therefore separated into segments of unequal illumination and may at first sight give the fictitious appearance of a dark cross. The eight achromatic brushes in this case separate the bright and dark segments, and are not very conspicuous.

112. Relation between wave velocities. If equation (4 a), Chapter VIII be solved with respect to v^2, the result is expressed in a manner which is difficult to interpret. A solution may be deduced independently in a more convenient form if the direction of the wave normal is defined by the angles included between it and the two optic axes. It can be proved that the length of an axis of a central section of a quadric completely determines the sum or difference of the angles which its plane makes with the two circular sections*. Whether it is the sum or difference depends on whether the selected axis is shorter or longer than the diameter of the circular section. If we therefore find the required relation for a principal axis in any one section, the same relation must hold for a principal axis of the same length in any other section.

Let the plane of the paper be the section containing the smallest and largest diameter AA' and CC' of the ellipsoid. If OM and OL (Fig. 148 a) be equal to the length of the intermediate principal axis of the ellipsoid, the circular sections are contained in the planes drawn at

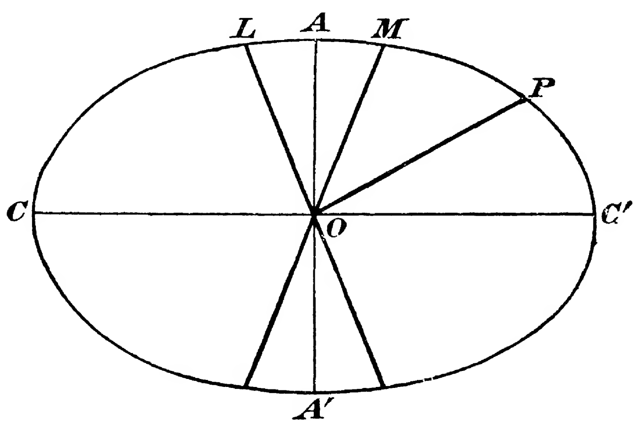


Fig. 148 a.

right angles to the plane of the paper and through these lines. The equation of the ellipse is $a^2x^2 + c^2y^2 = 1$. Any radius vector OP of this ellipse is a principal axis of the section passing through PQ and at right angles to the plane of the paper, as there is symmetry above and below that plane.

If v be the reciprocal of the length of OP forming an angle α with the minor axis,

$$v^2 = a^2 \sin^2 \alpha + c^2 \cos^2 \alpha \quad\text{.......................(2).}$$

As L and M must be equidistant from A it follows that if the angle POL and POM be ϕ' and ϕ respectively,

$$\alpha = \tfrac{1}{2}(\phi' \pm \phi),$$

where the lower sign is to be taken if OP is smaller than OM, so that P lies between M and L. By substitution we obtain

$$v^2 = a^2 \sin^2 \tfrac{1}{2}(\phi' \pm \phi) + c^2 \cos^2 \tfrac{1}{2}(\phi' \pm \phi) \quad\text{...........(3).}$$

* Salmon, *Geometry of Three Dimensions*, Art. 245.

The above equation must hold for any central section which has a principal diameter equal to v^{-1}. But each section has two axes having lengths which we may denote by $2v_1^{-1}$ and $2v_2^{-2}$. As the equation (3) must also include this second diameter, it follows that if the upper sign holds for one, the lower sign should be taken for the other. After an obvious trigonometrical transformation (3) is now replaced by

$$\left.\begin{aligned} 2v_1^2 &= (a^2+c^2) - (a^2-c^2)\cos(\phi'+\phi) \\ 2v_2^2 &= (a^2+c^2) - (a^2-c^2)\cos(\phi'-\phi) \end{aligned}\right\} \quad \text{............(4)},$$

and hence
$$\left.\begin{aligned} v_1^2 - v_2^2 &= (a^2-c^2)\sin\phi\sin\phi' \\ v_1^2 + v_2^2 &= (a^2+c^2) - (a^2-c^2)\cos\phi'\cos\phi \end{aligned}\right\} \quad \text{.........(5)}.$$
and

We may identify v_1 and v_2 with the velocities of plane waves in doubly refracting media, the wave normal forming angles ϕ and ϕ' with the optic axes. These quantities are therefore the two positive roots of equation (4 a), Chapter VIII.

In the particular case considered the wave-normal OP is situated in the plane containing the optic axes. The intersection of the wave-front with the ellipsoid of elasticity then passes through the axis of y and one of the wave-velocities must be equal to the reciprocal of the intermediate semi-axis. This can be verified as follows:

Write the second of equation (4) in the form

$$v_2^2 = a^2 - (a^2-c^2)\cos^2\tfrac{1}{2}(\phi-\phi').$$

Under the specified conditions the angle between the axis of x and each of the optic axes is equal to $\frac{1}{2}(\phi-\phi')$ and substituting the cosine of that angle as given in Art. 89 we find $v_2 = b$

If the difference between a and c is so small that its square may be neglected, we may write

$$\frac{1}{v_2} - \frac{1}{v_1} = \frac{v_1 - v_2}{v_1 v_2} = \frac{v_1^2 - v_2^2}{2v^3},$$

where v stands for the velocity to which both v_1 and v_2 approach when $a-c$ vanishes. For v we may therefore write either $\sqrt{ac}$ or $\frac{1}{2}(a+c)$, and for $2v^3$ we may write $ac(a+c)$.

Introducing the values of $v_1^2 - v_2^2$ from (4) we obtain

$$\frac{1}{v_2} - \frac{1}{v_1} = \frac{(a^2-c^2)}{(a+c)ac}\sin\theta_1\sin\theta_2,$$

or
$$\frac{1}{v_2} - \frac{1}{v_1} = \left(\frac{1}{c} - \frac{1}{a}\right)\sin\theta_1\sin\theta_2 \quad \text{....................(6)}.$$

113. Relation between ray velocities. The proposition contained in the last article represents a theorem which may be applied to any ellipsoid of semiaxes $1/a$, $1/b$, $1/c$; if v_1, v_2 are the reciprocals of the principal axes of a section which forms angles θ_1 and θ_2 with the circular sections. We may therefore write down at once the corresponding equations for the reciprocal ellipsoid, substituting the ray velocities s_1 and s_2 for $1/v_1$ and $1/v_2$. We obtain in this way

$$2s_1^{-2} = (a^{-2} + c^{-2}) - (a^{-2} - c^{-2}) \cos(\eta_1 + \eta_2),$$
$$2s_2^{-2} = (a^{-2} + c^{-2}) - (a^{-2} - c^{-2}) \cos(\eta_1 - \eta_2),$$
$$s_1^{-2} + s_2^{-2} = (a^{-2} + c^{-2}) - (a^{-2} - c^{-2}) \cos\eta_1 \cos\eta_2,$$
$$s_1^{-2} - s_2^{-2} = (a^{-2} - c^{-2}) \sin\eta_1 \sin\eta_2,$$

where η_1 and η_2 are the angles formed between the normal to the section and the axes of single ray velocities.

114. The surface of equal phase difference, or Isochromatic Surface. If we imagine a number of plane waves crossing at a point O (Fig. 149) in a crystalline medium, there being two wave velocities in each direction, we may construct a surface such that at any point P, belonging to this surface, such as P_1, P_2, the phase difference δ between the two wave-fronts which have OP for wave normals is the same. If ρ be the radius vector, v_1 and v_2 the wave velocities, the two optical distances from O to P are ρ/v_2 and ρ/v_1, hence the required surface has for equation:

$$\rho\left(\frac{1}{v_2} - \frac{1}{v_1}\right) = \text{constant}.$$

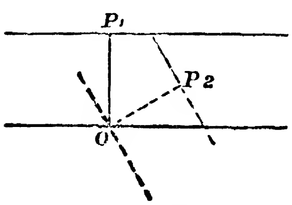


Fig. 149.

It will be sufficient to confine the discussion to the case of a small difference between the two wave velocities. We shall consider therefore $a - c$ and a fortiori $a - b$ to be so small that their squares may be neglected. We may then apply equation (6) and by introducing the principal indices of refraction $\mu_1 = \frac{1}{a}$, $\mu_3 = \frac{1}{c}$, the equation to the surface of equal phase difference is obtained in the form

$$\rho(\mu_3 - \mu_1) \sin\theta_1 \sin\theta_2 = \delta \quad\text{.....................(7).}$$

Unless highly homogeneous light is used, δ must not exceed a small multiple of a wave-length, if interference effects are to be observed. It follows that unless the observations are carried out close to one of the optic axes, in which case either $\sin\theta_1$ or $\sin\theta_2$ is small, $\mu_3 - \mu_1$ must be small. This justifies the simplification we have introduced in treating $a - c$ as a small quantity.

In uniaxal crystals, there is only one axis, so that putting $\theta_1 = \theta_2 = \theta$, the polar equation to the surface of equal phase difference or "isochromatic" surface then becomes

$$\rho(\mu_e - \mu_0) \sin^2\theta = \delta \quad\text{..........................(8).}$$

This surface is formed by the revolution about the optic axis of a family of curves for which the polar equation is represented by

(8) and which is drawn to scale in Fig. 150. Only half of the curves is shown, there being symmetrical halves below the line PQ. The scale is such that if the substance is Iceland Spar, and the length

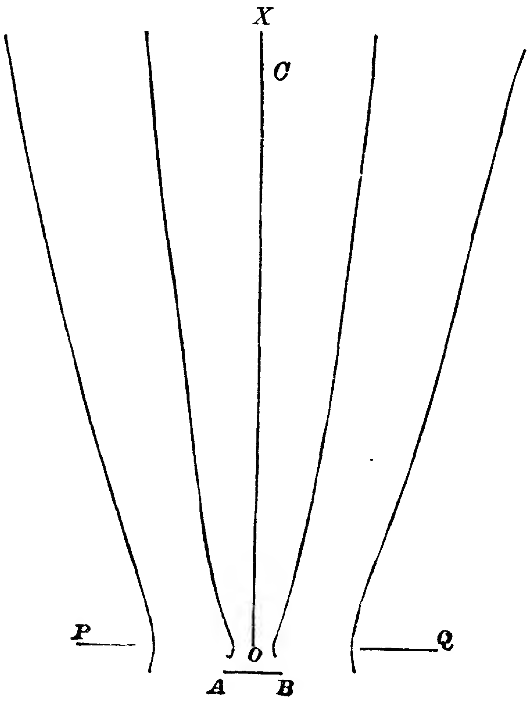


Fig. 150.

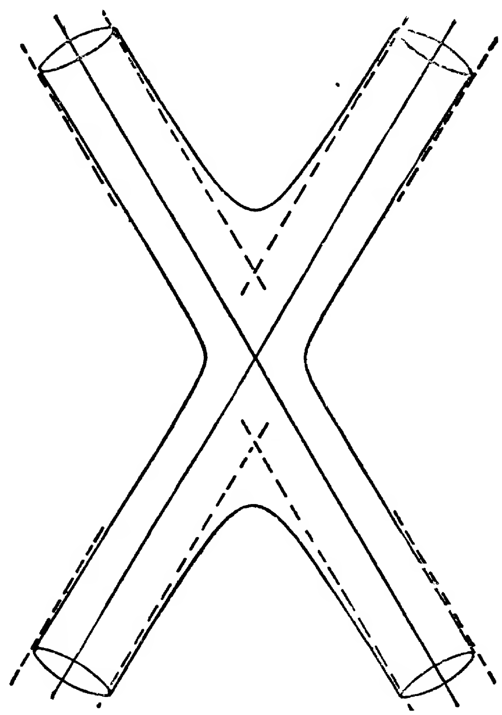


Fig. 151.

marked AB represents one millimetre, the inner curve is the isochromatic surface of phase difference equal to 100 wave-lengths, the wave-length being that of sodium light; the phase difference belonging to the outer curve is five times as great. OC is the optic axis. The upper portions of the curves are sensibly parabolic, because when θ is small, the radius vector ρ is nearly equal to its projection x on the optic axis, so that the equation to the curve becomes

$$z^2/x = \text{constant}.$$

In biaxal crystals the isochromatic surface has four sheets surrounding the optic axes. Their intersection with the plane containing these axes is represented in Fig. 151 for the case where the angle between the optic axes is 60°. When ρ is infinitely large, it follows from (7) that either θ_1 or θ_2 is zero. If θ_1 vanishes, θ_2 must be equal to σ, the angle included between the optic axes. For large values of ρ we may still take approximately $\theta_2 = \sigma$ and the equation to the isochromatic surface approaches therefore a surface the equation to which is by (7)

$$\rho \sin \theta_1 = \delta \operatorname{cosec} \sigma/(\mu_3 - \mu_1).$$

This is the equation to a circular cylinder, having one of the optic axes as axis. The intersection of this cylinder with the plane of the paper gives two straight parallel lines, which are the asymptotes to the curve which forms the intersection of the isochromatic surface with the plane containing the optic axes. If ρ' be the distance of the asymptotes from the origin

$$\rho' = \delta \operatorname{cosec} \sigma/(\mu_3 - \mu_1),$$

there are two similar asymptotes parallel to the second optic axis. These asymptotes are shown by dotted lines in the figure, and it will be noticed that each of them intersects one branch of the curve to which it is a tangent at infinity.

The two distances ρ_0 and ρ_1 of the vertices of the surface may be found by substituting $\theta_1 = \theta_2 = \frac{1}{2}\sigma$ and $\theta_1 = \frac{\pi - \sigma}{2}$, $\theta_2 = \frac{\pi + \sigma}{2}$ respectively. We then find

$$\rho_0 = \delta \operatorname{cosec}^2 \tfrac{1}{2}\sigma/(\mu_3 - \mu_1),$$
$$\rho_1 = \delta \sec^2 \tfrac{1}{2}\sigma/(\mu_3 - \mu_1).$$

115. Application of the Isochromatic Surface to the study of polarization. Let a doubly refracting plate, Fig. 152, receive light at different inclinations. An eye placed at E and looking towards a point S on the plate observes certain interference effects. Tracing the disturbance backwards from E, there will be two wave normals within the plate corresponding to SE. Let $OS = \rho$ be that wave normal which forms the smaller angle with OM the normal to the plate. According to Art 104 the difference in path at S, and therefore at E, of the two waves which have traversed the crystal is $\rho\left(\frac{1}{v_1} - \frac{1}{v_2}\right)$. A similar reasoning applies to the interference observed in the direction ET, O_1T being the direction of the wave normal inside the crystal. Draw OS_1 parallel to O_1T and EN at right angles to the plate. The interference seen at T is the same as that due to the phase difference at S_1 for waves propagated through O.

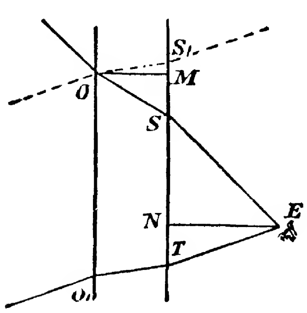


Fig. 152.

If i be the inclination of the line of sight

$$\tan i = \frac{NS}{NE}.$$

If r is the inclination of OS to OM

$$\tan r = \frac{MS}{MO},$$

$$\therefore \frac{NS}{MS} = \frac{NE}{MO} \frac{\tan i}{\tan r}.$$

For small values of i the ratio $\tan i/\tan r$ is nearly constant even in doubly refracting crystals. Representing this ratio by μ:

$$\frac{NS}{MS} = \mu \frac{NE}{MO},$$

and similarly

$$\frac{NT}{MS_1} = \mu \frac{NE}{MO}.$$

If an isochromatic surface be constructed with O as centre, it follows that its intersection with the upper surface of the plate enlarged in the ratio $\mu NE/MO$ gives the interference pattern as it is seen projected on the plate by an eye placed at E. When $\tan i/\tan r$ is not constant, there is a certain distortion due to the variability of that factor, but the general appearance remains unaltered.

As an example we may use Fig. 150 to construct the isochromatic lines for a plate of Iceland spar. Place the plate with its normal in the plane of the paper, its lower surface passing through O with OC along the optic axis. The upper surface will intersect the plane of the paper in a line which is at a distance from O equal to the thickness of the plate, the length of AB representing one millimetre. The intersection of the isochromatic surfaces which are formed by the revolution of the curves drawn in the figure about OC and the upper surface of the plate, will show the isochromatic lines for a phase difference of 100 and 500 wave-lengths. As all isochromatic surfaces may be obtained from one by increasing the length of the radius vector in a given proportion, we may obtain all isochromatic curves from the same surface by changing the scale. Thus to obtain the curve for which the retardation is ten wave-lengths, in the above example, we must, taking the inner curve, alter the scale, so that AB represents ·1 mm. In simple cases, this method of forming a rapid idea of the shape of the interference curves is very serviceable, the different curves being obtained by drawing the upper surface of the plate at different distances from the origin.

116. Isochromatic curves in uniaxal crystals. To study the intersection of the isochromatic surface and a plane drawn in any direction, construct a spherical triangle SNO (Fig. 153), such that if C be the centre of the sphere, CN is parallel to the normal of the plate, CO parallel to the optic axis, and CS parallel to any wave normal inside the plate. Also let

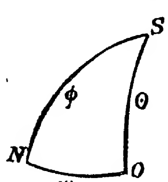


Fig. 153.

θ = angle between CS and CO,

ϕ = ,, ,, CN and CS,

ψ = ,, ,, CN and CO,

Δ = ,, ,, planes CNS and CNO.

In the spherical triangle NOS

$$\cos\theta = \cos\psi\cos\phi + \sin\psi\sin\phi\cos A\,;$$

$$\therefore\ \sin^2\theta = \sin^2\psi + \sin^2\phi\,(\cos^2\psi - \sin^2\psi\cos^2 A) - \sin 2\psi\sin\phi\cos\phi\cos A.$$

To obtain the isochromatic curves we must take the intersection between the surfaces given by (8):

$$\rho\sin^2\theta = \text{constant}$$

and the plane at which $\rho\cos\phi = e,$
where e is the thickness of the plate.

Eliminating ρ we obtain an equation for the curves in the form

$$\frac{\sin^2\theta}{\cos\phi} = \text{constant} \qquad (9).$$

We shall consider the angle of internal incidence to be so small that we may write sensibly $\dfrac{1}{\cos\phi} = 1 + \frac{1}{2}\sin^2\phi$, and rejecting all terms involving a higher power of ϕ than the second:

$$\frac{\sin^2\theta}{\cos\phi} = \sin^2\psi + \sin^2\phi\,(\tfrac{1}{2} + \tfrac{1}{2}\cos^2\psi - \sin^2\psi\cos^2 A) - \sin 2\psi\sin\phi\cos A.$$

An important special case occurs when the plate is cut parallel to the axis. In that case $\sin\psi = 1$ and $\sin 2\psi = 0$ so that the condition for equality of phase difference at the upper surface becomes

$$\sin^2\phi\,(\tfrac{1}{2} + \tfrac{1}{2}\cos^2\psi - \sin^2\psi\cos^2 A) = \text{constant},$$

or introducing the value of ψ

$$\sin^2\phi\,(\sin^2 A - \cos^2 A) = \text{constant}.$$

If we introduce rectangular coordinates with the pole of the plate normal N as centre, so that

$$x = e\sin\phi\sin A$$
$$y = e\sin\phi\cos A,$$

the equation to isochromatic curves reduces to

$$x^2 - y^2 = \text{constant}.$$

These curves are therefore rectangular hyperbolas, one of the axes being parallel to the direction of the optic axis and the other at right angles to it. If $\sin 2\psi$ does not vanish, then for small values of ϕ the term involving the first power of ϕ is the important one. Close to the normal therefore in a plate cut obliquely to the axis, the isochromatic lines are given by

$$\sin\phi\cos A = \text{constant},$$

which represents straight lines at right angles to the plane containing the optic axis and normal. When ϕ becomes sufficiently large for the second order terms to become appreciable, these lines become curved, but both terms together still represent conic sections.

Unless the normal to the plate is nearly coincident with the optic axis, there are no achromatic lines, as the axes of x and y remain sensibly parallel throughout the field.

117. Isochromatic Curves in Biaxal Crystals. We shall not follow out in detail the calculation in this case, but only indicate the method which may conveniently be adopted. Construct the spherical triangle NSO_1, NSO_2, Fig. 154, corresponding to NSO, Fig. 153, only with the difference that we have now two optic axes CO_1 and CO_2. Let A represent the angle between the sides NS and NG, where NG is a large circle bisecting the angle between NO_1 and NO_2. If ω be half the angle between the sides NO_2 and NO_1, we have in the triangles NO_2S and NO_1S

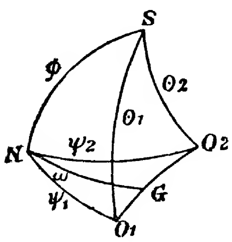


Fig. 154.

$$\cos\theta_2 = \cos\psi_2 \cos\phi + \sin\psi_2 \sin\phi \cos(A - \omega),$$

$$\cos\theta_1 = \cos\psi_1 \cos\phi + \sin\psi_1 \sin\phi \cos(A + \omega).$$

The angle 2ω may be obtained from ψ_1, ψ_2, and O_1O_2, the angle between the optic axes. From the above two equations we may obtain

$$\sin^2\theta_1 \sin^2\theta_2/\cos^2\phi,$$

expressed in a series proceeding by ascending powers of $\sin\phi$.

The isochromatic curves are found by combining (7) with $\rho\cos\phi = e$, and thus determined by putting $\sin\theta_1$, $\sin\theta_2/\cos\phi$ equal to a constant.

It is found that when the normal of the plate coincides with one of the axes of elasticity, the factor of the first and third powers of $\sin\phi$ is zero, and, neglecting $\sin^4\phi$, the condition for the isochromatic lines is obtained by putting the factor of $\sin^2\phi$ equal to zero. We thus obtain, as in the last article, the equation of rectangular hyperbolas. When the plate is cut obliquely the factor of $\sin\phi$ is the important one, and the curves close to the normal are straight lines, as with uniaxal crystals*.

118. Biaxal Crystals cut at right angles to the bisector of the angle between the optic axes. This case has special interest, and may be treated in a very simple manner, if the angle between the optic axes is small. Let OM_1, OM_2 be the directions of

* For the details of working out the general case, see Kirchhoff, *Vorlesungen über mathematische Optik*, p. 256.

the optic axes. When these nearly coincide with the normal, the angles OM_1P and OM_2P are nearly right angles, so that approximately,

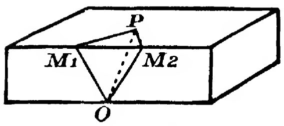


Fig. 155.

$$\sin\theta_1 = \frac{PM_1}{OP}; \ \sin\theta_2 = \frac{PM_2}{OP}.$$

Hence the equation to the isochromatic curve is

$$\frac{PM_1 \times PM_2}{OP} = \text{constant}.$$

If further, OP form a small angle with the normal, we may consider it to be constant and equal to the thickness of the plate. The isochromatic lines are in that case the lines on the surface of the plate which satisfy the equation

$$r_1 r_2 = \text{constant},$$

where r_1 and r_2 are measured from the points M_1 and M_2 on the crystal, such that plane waves traced back along the lines of vision EM_1 and EM_2 (Fig. 156) are refracted with their wave-normals parallel to the optic axes. The curves are so-called lemniscates. For small values of the constants they split up into separate curves, each surrounding one of the points M_1 or M_2. For large values of the constants, they are nearly circular, with the point halfway between M_1 and M_2 as centre. Figs. 4 and 5, Plate II, show the appearance.

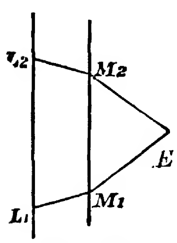


Fig. 156.

119. The Achromatic Lines in Biaxal Crystals. To trace the achromatic lines in a biaxal crystal cut so that the surface of the plate forms equal angles with the optic axes, we must introduce the condition that $\sin 2\alpha$ or $\sin 2\beta$ is zero. We begin by finding the locus of points on the surface of the plate (Fig. 155) at which the product $\sin\alpha \sin\beta$ has a constant value. The polarizer and analyser being fixed, α and β can only depend on the directions of the two vibrations emerging at the point P, and these are for the two rays respectively the external and internal bisectors of PM_1, PM_2. The problem therefore consists in finding the locus of a point P (Fig. 157) such that the bisector PK has a given direction which we take to be axis of Y, the origin of coordinates being the point halfway between the fixed points M_1 and M_2. If M_2H and M_1K be drawn at right angles to PK_1 the triangles PHM_2 and PKM_1 are similar. Hence

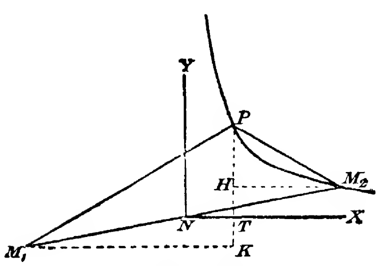


Fig. 157.

$$PK : PH = M_1K : M_2H,$$

and if x, y be the coordinates of P, and $\pm\xi$, $\pm\eta$ those of M_1 and M_2,

$$(y+\eta) : (y-\eta) = x + \xi : (\xi - x).$$

PLATE II.

FIG. 1.

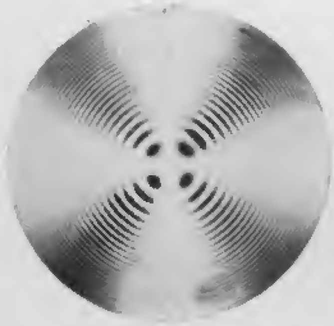


FIG. 2.

FIG. 3.

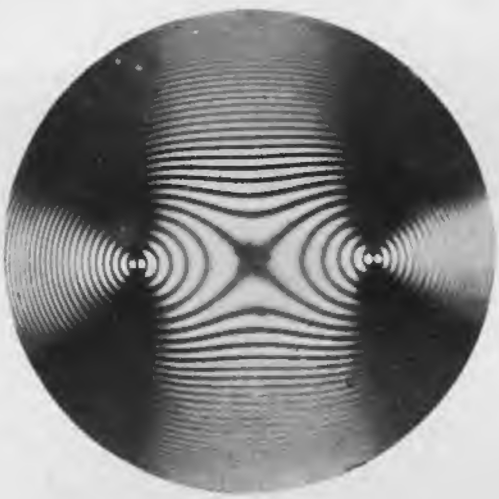


FIG. 4.

FIG. 5.

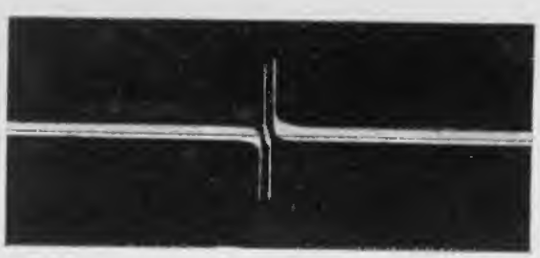


FIG. 6.

[To face page 210.

A simple transformation gives $xy = \xi\eta$. The locus of P is therefore a rectangular parabola passing through the fixed points M_1, M_2. As the choice of the axis of Y is arbitrary depending on the product $\sin\alpha \sin\beta$, we obtain a number of similar curves all passing through M_1 and M_2.

With crossed Nicols the intensity of the transmitted light is proportional to $\sin^2 2\alpha \sin^2 \delta/2$ (Art. 108). If α is zero there is no light at all, and the line along which there is this total extinction has just been shown to be a hyperbola. With small values of α, either positive or negative, $\sin 2\alpha$ is still very small, so that to either side of the line there is little illumination and a dark space appears, formed by a number of overlapping hyperbolas which are obtained by rotating the central curve, which is more particularly called the achromatic line about the points M_1 and M_2. The brush-like appearance of the dark space is shown in Plate II, Figs. 4 and 5. The asymptotes lie in the principal directions of the polarizer and analyser, and pass through the points halfway between M_1 and M_2. If the plate be rotated, the analyser and polarizer remaining fixed, the asymptotes remain in their places while the eccentricity of the hyperbola alters, the achromatic lines however still pass through the points M_1 and M_2.

If ρ be the distance of the central point from M_1 and θ the angle between ρ and the axis of x, $\xi\eta = \rho^2 \cos\theta \sin\theta$. On changing θ, the product has its maximum value when $\theta = 45°$. The eccentricity of the hyperbola has then its smallest value and M_1, M_2 are at the vertices. This is approximately the position in Fig. 4. If θ is zero or equal to a right angle, so that points M_1, M_2 lie on one of the coordinate axes, the central achromatic line coincides with the direction of the analyser or polarizer and the two brushes join together, and present an appearance similar to that of the cross in uniaxal crystals.

120. Measurement of angle between optic axes. The intersection of the isochromatic surface (Fig. 151) with planes drawn at different distances from O, shows that for small differences of path the interference rings surround the optic axes in closed curves. This affords a means of determining the angle between the optic axes. If a plate of a crystal cut symmetrically to the axes, as assumed in the last two articles, be mounted so that it can be rotated about an axis at right angles to the axes through an angle which can be measured, we may bring first one centre of the ring system belonging to one optic axis against a fixed mark in the observing telescope, and then the centre of the system belonging to the other axis. The angle of rotation is the so-called "apparent angle" between the optic axes, for it is clear that what is measured is the angle between the lines of vision M_1E and M_2E (Fig. 156). This angle is to be corrected for refraction to get the angles formed between L_1M_1 and L_2M_2.

121. Dispersion of Optic Axes. The position of the optic axes in crystal depends on the direction and the relative magnitudes of the axes of the ellipsoid of elasticity. As a rule the directions of the axes are fixed by the crystalline properties of the substance and do not depend on the wave-length. It is otherwise with the relative lengths of the axes. These are in the inverse ratio of the principal velocities v_1, v_2, v_3, and this ratio varies in some cases considerably on going from one colour to another. If it were necessary to trace accurately the isochromatic and achromatic lines, this so-called dispersion of the optic axes would have to be taken into account.

122. Two plates of a uniaxal crystal crossed. A great variety of effects may be produced by allowing light to traverse several plates in succession. We shall only consider one case, which is of some importance.

Let a plate be cut obliquely to the axis of a uniaxal crystal, and then divided into two halves which are therefore necessarily of the same thickness. Superpose the two halves and turn one of them through a right angle. We shall determine the shape of the isochromatic lines in this case.

The first plate produces a difference in optical length between two coincident wave normals, which as obtained from (9) is

$$\delta_1 = (\mu_0 - \mu_e)\frac{e \sin^2 \theta_1}{\cos \phi},$$

the meaning of the letters being the same as that of Article (116). The second plate being turned through a right angle, the direction of vibration in the ordinary and extraordinary rays is interchanged, so that the phase difference in that plate is

$$\delta_2 = (\mu_e - \mu_0)\frac{e \sin^2 \theta_2}{\cos \phi'}.$$

The values of cos ϕ and cos ϕ' are nearly equal for the double reason that ϕ is small, and that the difference between μ_e and μ_0 is small. Hence the total phase difference is proportional to

$$\sin^2 \theta_2 - \sin^2 \theta_1 = \cos^2 \theta_1 - \cos^2 \theta_2.$$

According to Art. 116

$$\cos \theta_1 = \cos \phi \cos \psi + \sin \phi \sin \psi \cos A.$$

To find the angle θ_2 which the optic axis makes with the plate normal in the upper plate, we have only to increase the angle A by a right angle, keeping all other quantities the same. Hence

$$\cos \theta_2 = \cos \phi \cos \psi - \sin \phi \sin \psi \sin A.$$

Neglecting higher powers of sin ϕ

$$\cos^2 \theta_1 - \cos^2 \theta_2 = \sin \phi \sin 2\psi (\cos A + \sin A).$$

Introducing rectangular coordinates, so that

$$e \sin \phi \cos A = x, \quad e \sin \phi \sin A = y,$$

the equation to the isochromatic line for which the total difference in optical length $\delta_1 + \delta_2$ is equal to $n\lambda$, becomes

$$(\mu_e - \mu_0) \sin 2\psi (x + y) = n\lambda \quad \text{.................. (10).}$$

This represents a series of parallel lines. The field of view is therefore crossed by a series of bands, the central one not being coloured. The bands are the wider apart the smaller ψ, so that if the bands are to be broad, the plate should be cut nearly normally to the optic axis. It is found that in this case, the departure from straightness which depends on terms involving $\sin^2 \phi$ is also small.

Two plates combined together in the manner described, form the essential portions of the "Savart" polariscope, which is the most delicate means we possess for detecting polarized light. The double plate is provided with an analyser, consisting of a Nicol prism or a Tourmaline plate. In both cases, the plane of transmittance through the analyser should bisect the angle between the principal planes of the Savart plates in order to get the most sensitive conditions. If the incident light be polarized at right angles to the plane of transmittance, the eye sees a dark central band accompanied on both sides by parallel coloured fringes. If the incident light be polarized parallel to the direction which can pass through the analyser, the central band is bright, and the whole effect is complementary to that observed in the previous case. By examining the light reflected from the sky or from almost any surface, the coloured fringes are noticed, and by rotating the whole apparatus we may find the direction in which the fringes are most brilliant and hence determine the plane of polarization of the incident light.

123. The Half Wave-length Plate. If plane polarized light falls normally on a plate of a crystal cut to such a thickness that the two waves are retarded relatively to each other by half a wave-length, or a multiple thereof, the transmitted beam is plane polarized. Let OX and OY be the two principal directions of vibration in the crystal, and α the angle between OX and the direction of vibration of the incident beam. The displacements resolved along OX and OY may then be expressed by

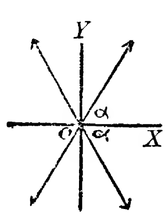


Fig. 158.

$$u = a \cos \alpha \cos \omega t,$$
$$v = a \sin \alpha \cos \omega t.$$

Then if the thickness of the plate be such that its optical length for the vibration along OY is half a wave-length greater, or half a wave-length less, than that for the vibration along OX, the displacements at emergence will be

$$u = a \cos \alpha \cos \omega t,$$
$$v = -a \sin \alpha \cos \omega t,$$

so that there is again plane polarization, but the angle of vibration forms an angle $-\alpha$ with the axis of x. The same holds for a retardation equal to any odd multiple of two right angles. For even multiples, the plane is that of the original vibration. These plates, in which a relative retardation of the two waves amounting to half a wave-length takes place, are called "Half Wave-length Plates" and are used in some instruments in which it is desired to fix the plane of polarization accurately. The simple Nicol does not permit of very exact adjustment, for while it is rotated slowly near the position of extinction, a broad dark patch is seen to travel across the field, and it is difficult to fix the exact position for which the centre of that patch is in the centre of the field of view. In the instruments in which a half-wave plate is used, that plate covers half the field of view. If ON and OM, Fig. 159, be the principal directions of the half-wave plate covering the left-hand portion of the field of view, and if the incident light vibrates along OP_1, the field of view will be divided by the plate into two portions, the directions of vibration at emergence being along OP_1, OP_2, equally inclined to ON. An eye examining the field through an analysing Nicol will find the two halves unequally illuminated, except where its principal plane is coincident with ON or at right angles to it. In the latter position, the luminosity of the field is small if α is small, and the eye is then very sensitive to small differences of illumination, so that the position of the analysing Nicol may be fixed with great accuracy. A half wave-length plate used in this fashion is the distinguishing feature of "Laurent's Polarimeter." The weak point of the arrangement lies in the effect of refrangibility on the retardation, in consequence of which a retardation of half a wave-length can only be obtained for a very limited part of the spectrum. Hence homogeneous light must be used with instruments which contain these plates.

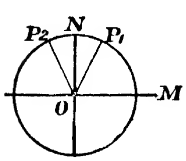


Fig. 159.

124. The Quarter Wave Plate. Plates in which the relative retardation of two waves is a quarter of a period, are called Quarter Wave Plates. They have the property of converting plane polarized light vibrating in a suitable direction into circularly polarized light. Let OX and OY be the two directions of vibration in the crystal, the vibration along OY being the one propagated most quickly.

Consider an incident plane polarized ray vibrating at an angle α to OX. The displacements in the incident vibrations are

$$u = a \cos \alpha \cos \omega t,$$
$$v = a \sin \alpha \cos \omega t.$$

At emergence the displacements may, by suitable adjustment of the origin of time, be expressed as

$$\left.\begin{array}{l} u = a \cos \alpha \cos \omega t \\ v = a \sin \alpha \cos (\omega t + \delta) \end{array}\right\} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (11).$$

In general this represents an elliptic vibration and we may investigate whether a point P moves clockwise or counter-clockwise through the ellipse. If α is in the first quadrant, then for $\omega t = \pi/2$, the x component of the displacement is zero, and the velocity in the x direction negative. Under these conditions, the rotation is positive (anti-clockwise) or negative (clockwise) according as the y displacement is positive or negative.

But under the above conditions at emergence for $\omega t = \pi/2$

$$v = - a \sin \alpha \sin \delta.$$

The rotation is positive or negative, therefore, according as $\sin \delta$ is negative or positive, hence if the total retardation is less than half a wave-length, the rotation is negative or clockwise. We should have got the opposite result if we had taken α to be in the second quadrant. Our conclusions may be formulated thus :—

If the retardation is less than half a wave-length, the rotation is from the direction OY, which belongs to the more quickly travelling wave, to the direction OP of the incident vibration, taking that branch of OP which forms an angle less than a right angle with OY.

If the retardation is between half a wave-length and a whole wave-length, the rotation is from the direction OP to the direction OY.

The displacements indicated by (11) when resolved along OP and at right angles to it, become

$$u' = a\left[\cos^2 \alpha \cos \omega t + \sin^2 \alpha \cos (\omega t + \delta)\right],$$
$$v' = a\left[\sin \alpha \cos \alpha \cos (\omega t + \delta) - \sin \alpha \cos \alpha \cos \omega t\right],$$

and if $\alpha = \dfrac{\pi}{4}$

$$u' = \tfrac{1}{2}a\left[\cos (\omega t + \delta) + \cos (nt)\right] = a \cos \tfrac{1}{2}\delta \cos (\omega t + \tfrac{1}{2}\delta),$$
$$v' = \tfrac{1}{2}a\left[\cos (\omega t + \delta) - \cos nt\right] \quad = - a \sin \tfrac{1}{2}\delta \sin (\omega t + \tfrac{1}{2}\delta),$$
$$\therefore \frac{u'^2}{\cos^2 \frac{1}{2}\delta} + \frac{v'^2}{\sin^2 \frac{1}{2}\delta} = a^2.$$

Hence the particle describes in general an ellipse having OP as one of its principal axes. When $\delta = \frac{\pi}{2}$ the ellipse becomes a circle. If therefore a plane wave be propagated through a doubly refracting substance, and if the incident vibration is equally inclined to OX and OY, then along the normal to the wave, the rays are plane, elliptically, or circularly polarized in regular succession. The state of vibration and the direction of rotation are indicated in Fig. 160 for equal distances from each other, each step in distance corresponding to a retardation of 45°.

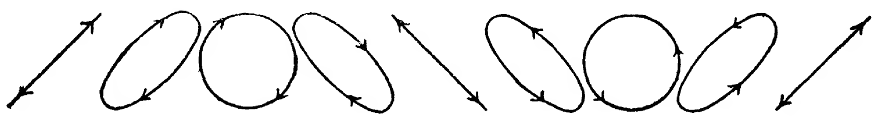


Fig. 160.

Thin plates of mica or gypsum may be obtained of the right thickness to give circular polarization. If the retardation is $3\lambda/4$ the effect is the same, but the rotation is left-handed in the same position relative to the crystal, where it was right-handed with a retardation of $\lambda/4$. We may call the direction OX the *axis* of the quarter plate, so that the direction of rotation for retardation of less than half a wave-length is from the direction of incident vibration to the direction of the axis, through the acute angle included between them. A retardation $\lambda/4 + n\lambda$ acts, so far as a particular wave-length is concerned, exactly like one of $\lambda/4$, but the difference in the refractive index for different colours has a more serious effect, the higher the value of n.

125. Application of Quarter Wave Plate. Besides being able to give, at any rate for one wave-length, light which is circularly polarized and rotating either in one direction or in the other, a quarter wave plate is useful for the investigation of elliptically polarized light. Elliptic polarization may always be represented by the superposition of two plane vibrations taking place in the direction of the axes of the ellipse and having a relative retardation of 90°. This phase difference is in one direction or another according as the elliptic path is right-handed or left-handed. A quarter wave plate with its axis parallel to one of the axes of the ellipse will increase or diminish the existing phase difference by another right angle, and the result is therefore plane polarization. If a and b are the semiaxes of the original ellipse, the direction of vibration after passing through the quarter wave plate will form an angle $\tan^{-1}\left(\pm\frac{b}{a}\right)$ with the direction along which a is measured, the $\pm$ sign being determined by the question whether the quarter plate increases or diminishes the original retardation.

126. Babinet's Compensator. This is an arrangement which has been successfully used for the study of elliptic polarization. It consists of two wedges of quartz, with their axes in the direction of the shading of the two surfaces in Figure 161. If a parallel beam of light traverses the system in the direction LN, the ray vibrating in the direction of the edge CD of the upper prism will pass through that upper prism more quickly, but through the lower prism more slowly, than the vibration at right angles to it. The central ray passes through equal thicknesses of both prisms. If plane polarized light which may be resolved parallel to AB and AC fall on the prism, the central ray will be plane polarized in the same direction as the incident light but on either side the rays will in general show elliptic polarization. At certain distances however the relative retardation of the two rays is two right angles and the transmitted ray will again be plane polarized. If the transmitted light be examined by a Nicol, properly placed, the field of view is seen to be traversed by parallel bands. If now the original light is elliptically polarized, the whole system of bands is the same as before but shifted sideways. In Babinet's Compensator, each of the wedges may be shifted parallel to itself, and in this way the central band may be brought back to its former position. The amount of displacement necessary to bring it back measures the relative retardation, and by its means the ratio of the axes of the ellipse may be determined.

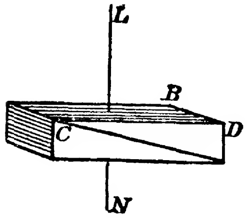


Fig. 161.

127. Circularly polarized light incident on a crystalline plate. We now consider the case where circularly polarized light falls on a crystalline plate and is then analysed by a Nicol prism or other plane polarizer. The incident light may be considered to be made of the superposition of two plane polarized waves having a relative retardation of a quarter of a wave-length. To fix our ideas, let the rotation of the incident light be anti-clockwise, the displacement along OX being represented by $a \cos \omega t$ and that along OY by $a \sin \omega t$. The direction of these axes may be chosen according to convenience and we may take them to be coincident with the principal directions of vibration inside the crystal. Let there be a retardation δ inside the crystal of that component which vibrates along OY. If the analyser is placed so that the light it can transmit vibrates along a direction forming an angle α with OX (the rotation from OX to OY being positive) the two parts of the beam leaving the analyser have amplitudes $a \cos \alpha$ and $a \sin \alpha$ and a phase difference of $\frac{1}{2}\pi + \delta$. Hence the intensity of the emergent light is

$$I = a^2 (1 - \sin 2\alpha \sin \delta) \qquad \text{(12)}.$$

This expression replaces equation (1) which holds when the incident light is plane polarized. The achromatic lines are determined by $\sin 2\alpha = 0$, and are therefore two lines at right angles to each other, parallel and perpendicular respectively to the principal plane of the analyser. The isochromatic lines are the curves for which δ is constant. If the plate is cut from a uniaxal crystal at right angles to the axis, the isochromatic lines are circles which in adjoining quadrants show complementary effects depending on the change of sign of $\sin 2\alpha$.

If the plate be examined by convergent or divergent light, the appearance, for positive values of δ, is that shown in Fig. 162, and for negative values of δ in Fig. 163. As the chromatic influence on the

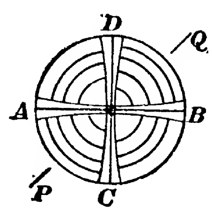


Fig. 162.

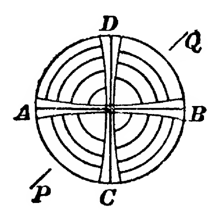


Fig. 163.

phase difference δ is the greater, the larger the phase difference, the first minimum observed with white light looks darker than the subsequent ones, the minima for the different colours overlapping more closely. We may refer to those two minima as the two dark spots, which lie in the first and third quadrants in Fig. 162 and in the second and fourth quadrants in Fig. 163.

The difference in the appearance gives us a useful criterion to distinguish between prolate and oblate crystals. Let it be required to study the intensity of light along the line NO (Fig. 148), which we take to be the axis of X. If OF be the axis of Y at O, and BC the direction of vibration transmitted by the analyser, α is in the first quadrant and $\sin 2\alpha$ in (12) is a positive quantity. OF being the direction of vibration of the ordinary ray, the retardation δ is positive for oblate crystals such as Iceland Spar, in which the ordinary ray is transmitted more slowly. Hence Fig. 158 represents the appearance. If the polarizer is placed at right angles to the analyser or along AB (Fig. 148), the axis of the quarter plate must, according to Art. 124, be placed parallel to AC if the rotation is to be anti-clockwise, as has been assumed. Hence

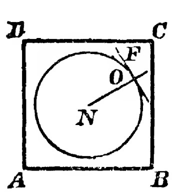


Fig. 148.

for oblate crystals, the line forming the dark spots is parallel to the axis of the quarter wave plate, while for prolate crystals (Fig. 163) the two lines are at right angles to each other. In both figures the line PQ marks the position of the axis of the quarter wave plate. It is easily seen that no difference is made in the appearance if the positions of the analyser and polarizer be interchanged. Hence the same rule holds whether the original rotation is clockwise or anti-clockwise, and we need only consider the relative positions of the axis of the quarter wave plate and the line joining the dark spots to decide between the two possible kinds of uniaxal crystals.

PART II.

CHAPTER X.

THEORIES OF LIGHT.

128. **Small strains in a small volume may always be treated as homogeneous strains.** Let α, β, γ represent the displacements within a strained body, and let the displacements be expressible as functions of the unstrained coordinates x, y, z of any point, so that

$$\alpha = f_1(x, y, z), \quad \beta = f_2(x, y, z), \quad \gamma = f_3(x, y, z).$$

Let further α', β', γ' be the displacements of a particle near x, y, z, which originally has coordinates $x+\xi$, $y+\eta$ and $z+\zeta$, then, neglecting squares of ξ, η, ζ, by Taylor's theorem:

$$\alpha' = \alpha + \frac{d\alpha}{dx}\xi + \frac{d\alpha}{dy}\eta + \frac{d\alpha}{dz}\zeta,$$

$$\beta' = \beta + \frac{d\beta}{dx}\xi + \frac{d\beta}{dy}\eta + \frac{d\beta}{dz}\zeta,$$

$$\gamma' = \gamma + \frac{d\gamma}{dx}\xi + \frac{d\gamma}{dy}\eta + \frac{d\gamma}{dz}\zeta,$$

ξ, η, ζ denoting the coordinates of the second particle relative to those of the first in the unstrained condition. If ξ', η', ζ' denote similarly the relative coordinates of the same two particles in the strained condition, we have

$$\xi' = (x + \xi + \alpha') - (x + \alpha) = \xi + \alpha' - \alpha,$$

or

$$\xi' = \left(1 + \frac{d\alpha}{dx}\right)\xi + \frac{d\alpha}{dy}\eta + \frac{d\alpha}{dz}\zeta$$

Similarly

$$\left.\begin{aligned} \xi' &= \left(1 + \frac{d\alpha}{dx}\right)\xi + \frac{d\alpha}{dy}\eta + \frac{d\alpha}{dz}\zeta \\ \eta' &= \frac{d\beta}{dx}\xi + \left(1 + \frac{d\beta}{dy}\right)\eta + \frac{d\beta}{dz}\zeta \\ \zeta' &= \frac{d\gamma}{dx}\xi + \frac{d\gamma}{dy}\eta + \left(1 + \frac{d\gamma}{dz}\right)\zeta \end{aligned}\right\} \quad \ldots\ldots\ldots\ldots\ldots(1).$$

These equations denote a homogeneous strain, for the linear relations between the strained and unstrained coordinates necessarily satisfy all the conditions laid down for such a strain by Thomson and Tait (*Natural Philosophy*, Vol. I. Art. 155): "*If when matter occupying any space is strained in any way, all pairs of points of its substance which are initially at equal distances from one another in parallel lines remain equidistant, it may be at an altered distance; and in parallel lines, altered it may be, from their initial direction; the strain is said to be homogeneous.*"

129. Simple Elongation. As a simple example of a homogeneous strain we may take the special case in which all coefficients except $\frac{d\alpha}{dx}$ vanish. This gives

$$\xi' = \left(1 + \frac{d\alpha}{dx}\right); \quad \eta' = \eta; \quad \zeta' = \zeta.$$

This is at once seen to represent a strain in which all lines parallel to OX are increased in the ratio $\left(1 + \frac{d\alpha}{dx}\right) : 1$, their distances from each other being unaltered. It is therefore a simple elongation along OX, the elongation being measured by $\frac{d\alpha}{dx}$. If $\frac{d\alpha}{dx}$ is small and if $\frac{d\beta}{dy}$ and $\frac{d\gamma}{dz}$ also have values which though small are not negligible, the strain consists of three small elongations along the three coordinate axes, superposed on each other. We denote these elongations by e, f, g, so that

$$e = \frac{d\alpha}{dx}, \quad f = \frac{d\beta}{dy}, \quad g = \frac{d\gamma}{dz}.$$

A cube having unit sides parallel to the coordinate axes, takes by the strain a volume equal to $(1+e)\cdot(1+f)\cdot(1+g)$, and neglecting small quantities of the second order, it is seen that the cubical dilatation which is the increase of volume of unit volume is measured by

$$e + f + g = \frac{d\alpha}{dx} + \frac{d\beta}{dy} + \frac{d\gamma}{dz} \quad \text{....................(2)}.$$

In a homogeneous strain all portions of a body have their volume increased or diminished in the same ratio, and we may therefore speak of the dilatation as a quantity belonging to the strain and independent of the position or shape of the portion of the body which we contemplate. This may formally be proved as follows:

Take three points having coordinates ξ_1, η_1, ζ_1; ξ_2, η_2, ζ_2; ξ_3, η_3, ζ_3 respectively.

The volume τ of the tetrahedron having these points as three of its vertices and the origin as the fourth, is equal to the sixth part of the determinant

$$\begin{vmatrix} \xi_1 & \eta_1 & \zeta_1 \\ \xi_2 & \eta_2 & \zeta_2 \\ \xi_3 & \eta_3 & \zeta_3 \end{vmatrix}.$$

This determinant is changed by the strain to

$$\begin{vmatrix} \xi_1' & \eta_1' & \zeta_1' \\ \xi_2' & \eta_2' & \zeta_2' \\ \xi_3' & \eta_3' & \zeta_3' \end{vmatrix}.$$

Substituting from equations (1) and applying a well-known theorem of determinants, it is found that the volume τ' of the strained tetrahedron is the sixth part of the product

$$\begin{vmatrix} \xi_1 & \eta_1 & \zeta_1 \\ \xi_2 & \eta_2 & \zeta_2 \\ \xi_3 & \eta_3 & \zeta_3 \end{vmatrix} \times \begin{vmatrix} 1+\dfrac{d\alpha}{dx} & \dfrac{d\alpha}{dy} & \dfrac{d\alpha}{dz} \\ \dfrac{d\beta}{dx} & 1+\dfrac{d\beta}{dy} & \dfrac{d\beta}{dz} \\ \dfrac{d\gamma}{dx} & \dfrac{d\gamma}{dy} & 1+\dfrac{d\gamma}{dz} \end{vmatrix}.$$

The second determinant simplifies, when the differential coefficients are so small that squares may be neglected, and becomes

$$1+\frac{d\alpha}{dx}+\frac{d\beta}{dy}+\frac{d\gamma}{dz}.$$

Hence
$$\tau'=\tau\left(1+\frac{d\alpha}{dx}+\frac{d\beta}{dy}+\frac{d\gamma}{dz}\right),$$

and
$$\frac{\tau'-\tau}{\tau}=\frac{d\alpha}{dx}+\frac{d\beta}{dy}+\frac{d\gamma}{dz}$$

measures the cubical dilatation.

130. Simple Shear. Consider a strain which is represented by the equations

$$\xi'=\xi+\frac{d\alpha}{dy}\eta,$$

$$\eta'=\frac{d\beta}{dx}\xi+\eta,$$

$$\zeta'=\zeta.$$

A point P on OX (Fig. 164), the axis along which both x and ξ are measured, keeps its x coordinate unchanged but is shifted parallel to OY through a distance $\xi d\beta/dx$, so that the line OX is turned through an angle $d\beta/dx$. Similarly a point Q on OY is shifted parallel to OX and the line OY is turned through an angle $d\alpha/dy$. The parallelogram $OP'RQ'$ has an area which, neglecting small quantities of the second order, is equal to $OP \times OQ$, so that the strain involves no sensible change of area, and as all z coordinates are unaltered, the strain involves no sensible change of volume. If the strained figure be rotated until OX' coincides with OX, it is seen that the total change may be represented as a sliding of all lines parallel to OX along themselves, the amount of the relative sliding being proportional to the distance between any two lines. The distance QQ' being $OQd\alpha/dy$, is increased by the rotation through an angle $d\beta/dx$ (bringing OP' into coincidence with OP), by an amount $OQd\beta/dx$, so that the sliding per unit distance is

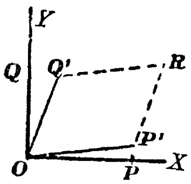


Fig. 164.

$$c = \frac{d\alpha}{dy} + \frac{d\beta}{dx}.$$

If the total strain is confined to such a sliding, it satisfies the condition of a simple shear (Thomson and Tait, § 171), c being the amount of the shear.

A simple shear may be produced by an elongation e in one direction, together with an equal contraction in a direction at right angles. Let OX and OY (Fig. 165) be the two directions. A length OA is changed by the strain to OA', where $OA' = (1+e)\,OA$. Take a point B on OY at a distance $OB = OA'$. If all lines along OY are reduced in the ratio $(1+e):1$, OB will be changed to OB', so that $OB' = OA$. If $OD = OA$, and $OC = OB$, the parallelogram $ABCD$ will be changed into $A'B'C'D'$. Imagine $A'B'C'D'$ to be transposed so that $A'B'$ is made to coincide with AB, and it will be seen that the whole change is equivalent to a sliding of the lines parallel to AB along their own lengths. If θ be the angle between AD and a line drawn at right angles to AB, the amount of sliding per unit distance is $2\tan\theta$.

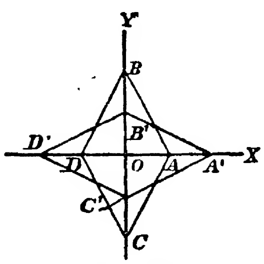


Fig. 165.

If further, α is the angle between OB and AB, $\theta + 2\alpha = \frac{1}{2}\pi$, so that the amount of sliding is $2\cot 2\alpha = \cot\alpha - \tan\alpha$.

Now $$\tan\alpha = \frac{OA}{OB} = \frac{OA}{OA'} = \frac{1}{1+e}$$

$$= 1 - e \text{ (approximately, if } e \text{ is a small quantity).}$$

Hence the amount of sliding is $2e$, neglecting small quantities of the second order.

To sum up: "A simple extension in one set of parallels, and a simple contraction of equal amount in any other set perpendicular to those, is the same as a simple shear in either of the two sets of planes cutting the two sets of parallels at 45°. And the numerical measure of this shear, or simple distortion, is equal to *double* the amount of the elongation or contraction (each measured of course per unit length)." (Thomson and Tait, § 681.)

131. Components of Strain. Neglecting small quantities of the second order, the strain represented by the equations (1) may be imagined to be produced by the superposition of six separate steps, which are three simple elongations and three simple shears. Beginning at first with the three elongations, the resulting change is represented by

$$\xi_1' = \left(1 + \frac{d\alpha}{dx}\right)\xi,$$

$$\eta_1' = \left(1 + \frac{d\beta}{dy}\right)\eta,$$

$$\zeta_1' = \left(1 + \frac{d\gamma}{dz}\right)\zeta.$$

We next suppose a change indicated by

$$\xi_2' = \xi_1' + \frac{d\alpha}{dy}\eta_1',$$

$$\eta_2' = \frac{d\beta}{dx}\xi_1' + \eta_1',$$

$$\zeta_2' = \zeta_1',$$

which according to the previous article is a simple shear of amount $c = \frac{d\alpha}{dy} + \frac{d\beta}{dx}$ in the plane of xy. By substitution we find, neglecting squares of small quantities, the total change so far to be given by

$$\xi_2' = \left(1 + \frac{d\alpha}{dx}\right)\xi + \frac{d\alpha}{dy}\eta,$$

$$\eta_2' = \frac{d\beta}{dx}\xi + \left(1 + \frac{d\beta}{dy}\right)\eta,$$

$$\zeta_2' = \left(1 + \frac{d\gamma}{dz}\right)\zeta.$$

If we further superpose shears of amount

$$a = \frac{d\beta}{dz} + \frac{d\gamma}{dy} \text{ in the plane of } yz$$

and $$b = \frac{d\gamma}{dx} + \frac{d\alpha}{dz} \text{ in the plane of } zx,$$

we return to the set of equations (1). The six quantities e, f, g, a, b, c, are called the components of the strain.

132. Homogeneous Stress. "When through any space in a body under the action of force, the mutual force between the portions of matter on the two sides of any plane area is equal and parallel to the mutual force across any equal, similar, and parallel plane area, the stress is said to be homogeneous through that space. In other words, the stress experienced by the matter is homogeneous through any space if all equal similar and similarly turned portions of matter within this space are similarly and equally influenced by force." (Thomson and Tait, § 659.)

Consider a unit cube (Fig. 166) subject to homogeneous internal stresses and in equilibrium. The stress on each of the six sides may be decomposed into three along the coordinate axes, but as, from the definition of a homogeneous stress, the forces acting in the same direction across opposite faces must be equal, we need only consider three faces of the cube. We denote by X_x, Y_x, Z_x, the three components of force acting on the face yz, the index x indicating that the face is normal to the axis of x. Similarly X_y, Y_y, Z_y, and X_z, Y_z, Z_z, indicate the components acting on the faces normal to the axes of y and z respectively. If we consider the force which acts on the cube from the outside, two stresses X_x act in opposite directions on the two faces normal to OX. If we take X_x to be positive the two forces tend to produce elongation. Similarly Y_y and Z_z are stresses tending to produce elongations along the axes of y and z respectively.

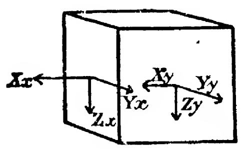


Fig. 166.

The force X_z (Fig. 167) is a tangential force acting in opposite directions on two opposite faces, but not along the same line, so that a couple of moment X_z is formed. We take X_z to be positive when, as drawn in the figure, the force acting on a surface parallel to xy from below is along the negative axis of y, the axis of z being positive upwards. But the two forces Z_x also form a couple, which however tends to produce rotation round OY in the opposite direction, hence for equilibrium, it follows that

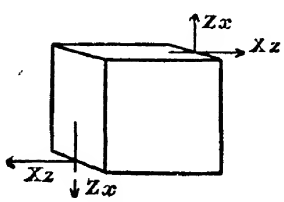


Fig. 167.

$$X_z = Z_x.$$

The two equal couples X_z and Z_x form together a simple shearing stress. It may be proved in the same manner that

$$Y_x = X_y,$$
$$Z_y = Y_z.$$

The six quantities

$$X_x,\ Y_y,\ Z_z;\ Y_z = Z_y;\ Z_x = X_z;\ X_y = Y_x,$$

completely define a homogeneous stress. We shall introduce the notation of Thomson and Tait, and write for these six components of stress

$$P, Q, R, S, T, U.$$

133. Shearing stress produced by combined tension and pressure at right angles. Let $ABCD$ be a section of a cube, which is subject to a uniform tension P at right angles to BC, and a uniform pressure at right angles to CD. No stress is supposed to act at right angles to the plane of the paper. Let H, K, L, M be the middle points of the sides of the square $ABCD$, and draw the square $HKLM$. If the part HBK is in equilibrium, a force must act on the plane which is at right angles to the plane of the paper, and passes through HK. The elementary laws of Statics show that this force must be in the plane, and that its value per unit surface is P. The rectangular volume of $HKLM$ is therefore acted on by tangential stresses of the nature of shearing stresses, or:

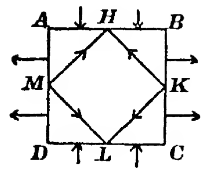


Fig. 168.

"A longitudinal traction (or negative pressure) parallel to one line and an equal longitudinal positive pressure parallel to any line at right angles to it, is equivalent to a shearing stress of tangential tractions parallel to the planes which cut those lines at 45°. And the numerical pressure of this shearing stress, being the amount of the tangential traction in either set of planes, is equal to the amount of the positive or negative normal pressure, *not* doubled." (Thomson and Tait, § 681.) The caution at the end of the quotation is necessitated by the fact that in the analogous proposition referring to shears, the amount of the shear is obtained by doubling the elongation, as has been proved in Art. 130.

134. Connexion between Strains and Stresses. If a simple shearing stress, as defined in Art. 132, act on a homogeneous body, it produces a shearing strain, and the ratio of the stress to the strain is the resistance to change of shape or the "Rigidity" of the substance. Calling the rigidity n, it follows that we may put

$$S = na\,;\; T = nb\,;\; U = nc \quad \text{.....................(3)}$$

in isotropic bodies.

The three stresses P, Q, R produce elongations e, f, g, and there must be a linear relationship between them. Also by symmetry a stress along OX must produce the same contraction in all directions at right angles to itself. Hence A and B being constants, we may write down at once the equations

$$\left.\begin{aligned} P &= Ae + B(f+g) \\ Q &= Af + B(g+e) \\ R &= Ag + B(e+f) \end{aligned}\right\} \quad \text{........................(4).}$$

It remains to prove how A and B are connected with the rigidity and the bulk modulus. If e, f, g are equal

$$P = Q = R = e\,(A + 2B).$$

Hence the stress is uniform.

But the cubical dilatation being $3e$ and the bulk modulus being equal to the ratio of the uniform stress P to the cubical dilatation, it follows that

$$3k = A + 2B \quad \text{.........................(5)}.$$

As a second special case take $R = 0$, and $Q = -P$, which conditions indicate a shearing stress in planes equally inclined to the axis of X and Y, and these will cause a shearing strain equal in amount to P/n. This shearing strain is equivalent by Art. 130 to an elongation in the direction of P of $P/2n$, and an equal contraction in the direction of Q. Substituting $e = -f = P/2n$ into the first of the equations (4), we find if $g = 0$

$$2n = A - B.$$

Combining this with (5), it follows that

$$A = k + \tfrac{4}{3}n, \quad B = k - \tfrac{2}{3}n \quad \text{....................(6)}.$$

In place of the components of strain, we may introduce their equivalents in terms of the displacement (Arts. 129 and 130). Equations (3) and (4) then become

$$S = n\left(\frac{d\beta}{dz} + \frac{d\gamma}{dy}\right), \quad T = n\left(\frac{d\gamma}{dx} + \frac{d\alpha}{dz}\right), \quad U = n\left(\frac{d\alpha}{dy} + \frac{d\beta}{dx}\right) \text{......(7)},$$

and

$$\left.\begin{aligned} P &= A\frac{d\alpha}{dx} + B\left(\frac{d\beta}{dy} + \frac{d\gamma}{dz}\right) \\ Q &= A\frac{d\beta}{dy} + B\left(\frac{d\gamma}{dz} + \frac{d\alpha}{dx}\right) \\ R &= A\frac{d\gamma}{dz} + B\left(\frac{d\alpha}{dx} + \frac{d\beta}{dy}\right) \end{aligned}\right\} \quad \text{.................(8)}.$$

135. Equations of Motion in a disturbed medium. Returning to the stresses acting on the cube in Art. 132, we consider the case where these stresses are not constant through the volume, but alter slowly from place to place. If the distance between the two faces of the cube which are at right angles to the axis of X is dx, there will be a force

$$X_x dy\, dz$$

acting in the negative direction on the face which is coincident with the coordinate plane and a force on the opposite face equal to

$$\left(X_x + \frac{dX_x}{dx}dx\right)dy\,dz.$$

These combine to a resultant

$$\frac{dX_x}{dx}\,dx\,dy\,dz.$$

Similarly the force $X_z dx dy$ acting on the plane xy in the direction of x together with the force

$$\left(X_z + \frac{dX_z}{dz}\,dz\right)dx\,dy$$

combine to a resultant

$$\frac{dX_z}{dz}\,dx\,dy\,dz,$$

and the forces in that same direction are complete when we have added the resultant

$$\frac{dX_y}{dy}\,dx\,dy\,dz$$

of the two forces which act on the faces which are normal to the axis of y. If ρ be the density of the substance, so that $\rho dxdydz$ be the mass of the volume considered, and if α be the displacement in the x direction, the equations of motion may be written down by the laws of dynamics, leaving out the factor $dxdydz$ on both sides,

$$\rho\frac{d^2\alpha}{dt^2} = \frac{dX_x}{dx} + \frac{dX_y}{dy} + \frac{dX_z}{dz}.$$

Similarly

$$\rho\frac{d^2\beta}{dt^2} = \frac{dY_x}{dx} + \frac{dY_y}{dy} + \frac{dY_z}{dz},$$

$$\rho\frac{d^2\gamma}{dt^2} = \frac{dZ_x}{dx} + \frac{dZ_y}{dy} + \frac{dZ_z}{dz}.$$

Re-introducing the notation of Thomson and Tait, the equations become

$$\rho\frac{d^2\alpha}{dt^2} = \frac{dP}{dx} + \frac{dU}{dy} + \frac{dT}{dz},$$

$$\rho\frac{d^2\beta}{dt^2} = \frac{dU}{dx} + \frac{dQ}{dy} + \frac{dS}{dz},$$

$$\rho\frac{d^2\gamma}{dt^2} = \frac{dT}{dx} + \frac{dS}{dy} + \frac{dR}{dz}.$$

To eliminate the stresses use equations (7) and (8).

Substituting the values of P, Q, A and B from (5), (6), (7) and (8), and rearranging the terms, we obtain

$$\left.\begin{aligned}
\rho\frac{d^2\alpha}{dt^2} &= n\left(\frac{d^2\alpha}{dx^2} + \frac{d^2\alpha}{dy^2} + \frac{d^2\alpha}{dz^2}\right) + (k + \tfrac{1}{3}n)\frac{d}{dx}\left(\frac{d\alpha}{dx} + \frac{d\beta}{dy} + \frac{d\gamma}{dz}\right)\\
\rho\frac{d^2\beta}{dt^2} &= n\left(\frac{d^2\beta}{dx^2} + \frac{d^2\beta}{dy^2} + \frac{d^2\beta}{dz^2}\right) + (k + \tfrac{1}{3}n)\frac{d}{dy}\left(\frac{d\alpha}{dx} + \frac{d\beta}{dy} + \frac{d\gamma}{dz}\right)\\
\rho\frac{d^2\gamma}{dt^2} &= n\left(\frac{d^2\gamma}{dx^2} + \frac{d^2\gamma}{dy^2} + \frac{d^2\gamma}{dz^2}\right) + (k + \tfrac{1}{3}n)\frac{d}{dz}\left(\frac{d\alpha}{dx} + \frac{d\beta}{dy} + \frac{d\gamma}{dz}\right)
\end{aligned}\right\}\ \ldots(9).$$

These equations govern wave propagation in all elastic media. We may obtain from them the characteristic equations for the longitudinal waves of sound by putting the rigidity n of the medium equal to zero. When applied to light, the medium is taken as incompressible, so that

$$\frac{d\alpha}{dx} + \frac{d\beta}{dy} + \frac{d\gamma}{dz} = 0,$$

but k at the same time becomes infinitely large. Writing

$$\delta = \frac{d\alpha}{dx} + \frac{d\beta}{dy} + \frac{d\gamma}{dz},$$

$$p = k\delta,$$

and

$$\nabla^2 = \frac{d^2}{dx^2} + \frac{d^2}{dy^2} + \frac{d^2}{dz^2},$$

the equations become

$$\left.\begin{aligned} \rho\frac{d^2\alpha}{dt^2} &= n\nabla^2\alpha + \frac{dp}{dx} \\ \rho\frac{d^2\beta}{dt^2} &= n\nabla^2\beta + \frac{dp}{dy} \\ \rho\frac{d^2\gamma}{dt^2} &= n\nabla^2\gamma + \frac{dp}{dz} \end{aligned}\right\} \quad \dots\dots\dots\dots\dots\dots\dots\dots(10).$$

These equations, together with certain relations which must hold at the surfaces of the elastic body, constitute the elastic solid theory of light.

For plane waves, the displacements are the same at all points of the wave-front, which we may imagine to be at right angles to the axis of z. The differential coefficient of α, β, γ with respect to x and y must therefore vanish. The equations (9) then reduce to

$$\rho\frac{d^2\alpha}{dt^2} = n\frac{d^2\alpha}{dz^2}; \quad \rho\frac{d^2\beta}{dt^2} = n\frac{d^2\beta}{dz^2}; \quad \rho\frac{d^2\gamma}{dt^2} = (k + \tfrac{4}{3}n)\frac{d^2\gamma}{dz^2} \dots\dots(11).$$

The last equation represents a longitudinal wave propagated with infinite velocity and having no relation to any observed phenomenon of light. Each of the first two equations represents a rectilinear wave propagated with velocity $\sqrt{n/\rho}$, a result already deduced by the simpler but less general methods of Art. 12.

The investigation of wave propagation in crystalline media presents great difficulties. The simplest hypothesis from a mathematical point of view is that of assuming that the inertia of the medium may differ for displacements in different directions. By substituting ρ_1, ρ_2, ρ_3, respectively, for ρ on the left-hand side of equations (9), we obtain equations which lead to a wave surface which is similar to, but not identical with, Fresnel's wave surface. A theory of double refraction based on this hypothesis was brought forward by Lord Rayleigh*, but abandoned because observations made by Stokes, and afterwards by Glazebrook, decided in favour of Fresnel's surface. Instead of taking the inertia as variable, we may adopt the very plausible hypothesis that the rigidity is different in different directions. Thus different values of n in the first two equations (11) would give two plane waves propagated with different velocities, along the axis of z. A general theory cannot however be formed by a simple modification of the equations holding for isotropic media. According to Green†, there may be twenty-one different coefficients defining the properties of crystalline media, which shows the complication we might be led into if we wished to attack the problem in its most general form.

136. Equations of the Electromagnetic Field. The line integral of the magnetic force round a closed curve is numerically equal to the electric current through the curve multiplied by 4π. It is shown in treatises on Electricity that the mathematical expression of this law is contained in the three equations:

$$\left.\begin{aligned} 4\pi u &= \frac{d\gamma}{dy} - \frac{d\beta}{dz} \\ 4\pi v &= \frac{d\alpha}{dz} - \frac{d\gamma}{dx} \\ 4\pi w &= \frac{d\beta}{dx} - \frac{d\alpha}{dy} \end{aligned}\right\} \quad \text{..........................(12)},$$

where α, β, γ are the components of magnetic force, and u, v, w the components of current density. The factor 4π depends on the units chosen, which are those of the electromagnetic system.

Another proposition which embodies Faraday's laws of electromagnetic induction states that if a closed curve encloses lines of magnetic induction which vary in intensity, an electromotive force acts round the curve, and the line integral of the electric force round the closed curve is equal to the rate of diminution of the total magnetic induction through the circuit. This leads to the equations

$$\left.\begin{aligned} -\mu\frac{d\alpha}{dt} &= \frac{dR}{dy} - \frac{dQ}{dz} \\ -\mu\frac{d\beta}{dt} &= \frac{dP}{dz} - \frac{dR}{dx} \\ -\mu\frac{d\gamma}{dt} &= \frac{dQ}{dx} - \frac{dP}{dy} \end{aligned}\right\} \quad \text{.....................(13)},$$

where $\mu\alpha$, $\mu\beta$, $\mu\gamma$ are the components of magnetic induction, μ being the permeability, and P, Q, R those of electric force.

* *Collected Works*, Vol. I. p. 111.

† *Collected Works*, p. 245.

The two sets of equations may be taken to represent experimental facts and to be quite independent of any theory. The equations would be equally true if we considered electric and magnetic forces to be due to action at a distance.

There are some additional equations to be considered.

Differentiating equations (12) with respect to x, y z respectively and adding, we find:

$$\frac{du}{dx}+\frac{dv}{dy}+\frac{dw}{dz}=0 \quad \text{......................(14)}.$$

Similarly we derive from (13), if μ be constant:

$$\frac{d}{dt}\left(\frac{d\alpha}{dx}+\frac{d\beta}{dy}+\frac{d\gamma}{dz}\right)=0.$$

This shows that the expression in brackets is constant as regards time and we know that in an unmagnetized medium it is zero. Hence we may write for our present purpose:

$$\frac{d\alpha}{dx}+\frac{d\beta}{dy}+\frac{d\gamma}{dz}=0 \quad \text{......................(15)}.$$

137. Maxwell's Theory. The fundamental principle of Maxwell's theory lies in his conception of the possibility of an electric current in dielectrics and the manner in which this current depends on electric force. His views are best explained by an analogy taken from the theory of stress and strain. A stress in an elastic solid produces a displacement: if the stresses increase, the displacements increase, and the change of displacements constitutes a transference of matter. Taking this as a guide we may imagine the medium to yield in some unknown manner to the application ef electric force, and if so, the rate of change of that force will cause a "flow" which according to Maxwell is identical in all its effects with an electric current.

If the electric force is E, the electric current is proportional to dE/dt, and if the law that the total flow is the same across all cross-sections of a circuit holds good for these so-called "displacement currents" or "polarization currents," it can be shown that the current is equal to $(4\pi)^{-1}KdE/dt$, where K is the specific inductive capacity of the medium. In a conductor, the current would, according to Ohm's law, be CE, where C is the conductivity. If we imagine a medium to possess both specific inductive capacity and conductivity, we must introduce an expression which includes both cases and put the current

$$\left(C+\frac{1}{4\pi}K\frac{d}{dt}\right)E \quad \text{......................(16)}.$$

Confining ourselves at present to non-conductors and resolving along the three coordinate axes, we have

$$u=\frac{1}{4\pi}K\frac{dP}{dt}\,;\ v=\frac{1}{4\pi}K\frac{dQ}{dt}\,;\ w=\frac{1}{4\pi}K\frac{dR}{dt}\ \ldots\ldots(17).$$

These equations allow us to combine (12) and (13) so as to obtain two fresh sets containing respectively only the magnetic and the electric forces.

138. Differential equation for propagation of electric and magnetic disturbances in dielectric media.

Equations (12) with the help of (17) become

$$\left.\begin{aligned} K\frac{dP}{dt}&=\frac{d\gamma}{dy}-\frac{d\beta}{dz}\\ K\frac{dQ}{dt}&=\frac{d\alpha}{dz}-\frac{d\gamma}{dx}\\ K\frac{dR}{dt}&=\frac{d\beta}{dx}-\frac{d\alpha}{dy}\end{aligned}\right\}\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(18).$$

Differentiate each of the equations (13) with respect to the time, eliminate P, Q, R, by means of (18), and use (15), when the following sets of equations, involving only magnetic forces, will be obtained :

$$K\mu\frac{d^2\alpha}{dt^2}=\nabla^2\alpha\,;\quad K\mu\frac{d^2\beta}{dt^2}=\nabla^2\beta\,;\quad K\mu=\frac{d^2\gamma}{dt^2}=\nabla^2\gamma\ldots\ldots(19).$$

We may eliminate the magnetic forces in a similar manner and obtain

$$K\mu\frac{d^2P}{dt^2}=\nabla^2P;\ K\mu\frac{d^2Q}{dt^2}=\nabla^2Q\,;\ K\mu\frac{d^2R}{dt^2}=\nabla^2R\ldots\ldots(20).$$

These equations show that the magnetic and electric forces are propagated with a velocity $1/\sqrt{K\mu}$. In the electromagnetic system of units, $\mu=1$ in vacuo, and differs very little from that value in any known dielectric. K the specific inductive capacity is, in vacuo, unity when the *electrostatic* system of units is employed, but in the electromagnetic system K is numerically equal to $1/v^2$, if v is equal to the number of electrostatic units of quantity which are contained in an electromagnetic unit. This number, which gives the velocity of propagation of electromagnetic waves in vacuo, may be determined by experiment, and is found, within the errors of experiment, to be equal to the velocity of light in vacuo. Both velocities differ from 3×10^{10} probably by not more than one part in a thousand.

Maxwell's theory, which is embodied in equations (19) or (20), leads therefore to the remarkable conclusion that an electromagnetic disturbance is propagated with a finite velocity which is equal to the velocity of light. This conclusion has been amply verified by the celebrated experiments of Hertz. Kirchhoff* had already in 1857 pointed out that a longitudinal electric disturbance is propagated in a wire with a velocity equal to that of light, but it was left to Maxwell to discover the reason for this coincidence.

If both the disturbance of light and the electromagnetic wave are propagated through the same medium with the same velocity, the conclusion is irresistible that both phenomena are identical in character. This conclusion constitutes the so-called "Electromagnetic Theory of Light." The electromagnetic theory of light establishes for the propagation of a luminous disturbance, equations which in several instances, as will appear, fit the facts better than the older elastic solid theory, but it should not be forgotten that it furnishes no explanation of the nature of light. It only expresses one unknown quantity (light) in terms of other unknown quantities (magnetic and electric disturbances), but magnetic and electric stresses are capable of experimental investigation, while the elastic properties of the medium through which, according to the older theory, light was propagated, could only be surmised from the supposed analogy with the elastic properties of material media. Hence it is not surprising that the electromagnetic equations more correctly represent the actual phenomena. Whatever changes be introduced in future, in our ideas of the nature of light, the one great achievement of Maxwell, the proof of the identity of luminous and electromagnetic disturbances, will never be overthrown.

139. Refraction. We have so far only considered the propagation of waves in vacuo. According to equations (20), the squares of the velocities of propagation in two media having identical magnetic permeabilities, ought to be inversely as their specific inductive capacities. If therefore K_0 be the inductive capacity of the vacuum, K_1 that of any dielectric, the "refractive index" ought to be equal to $\sqrt{K_1/K_0}$. This relation is approximately verified in the case of a few gases, as shown in the following table, which contains the square roots of specific inductive capacities (D) as measured by Klemencic†, and the refractive indices (n) of the same gases for Sodium light, as measured by G. W. Walker‡. Both constants are reduced to a temperature of 0° C., and a pressure of 760 mm.

Nature of Gas.	(D)	(n)
Air	1·000293	1·000293
Hydrogen	1·000132	1·000141
Carbon dioxide	1·000492	1·000451
Sulphur dioxide	1·000477	1·000676

* *Pogg. Ann.* Vol. c. p. 193 (1857).

† *Wien Ber.* (2) Vol. xci. p. 1 (1885).

‡ *Trans. Roy. Soc.* A. Vol. cci. p. 435 (1903).

The discrepancy for sulphur dioxide is already well marked.

For solids and liquids the relation altogether fails. Thus water has a specific inductive capacity which is 80 times greater than that of air, and its refractive index should therefore be equal to 9, or six times larger than its actual value. But these discrepancies are not surprising, for we have left a factor out of consideration, which to a great extent dominates the phenomenon of refraction, and that is absorption. The theoretical relationship really applies only to waves of infinite length, but in most cases we know nothing of the refractive index for very long waves. The subject will be further discussed in the next Chapter.

140. Direction of Electric and Magnetic Forces at right angles to each other. If we confine ourselves for the sake of simplicity to waves, parallel to the plane of xy, we must take in equations (13) and (17) all quantities to be independent of x and y: these equations then become

$$\left.\begin{aligned} \mu\frac{d\alpha}{dt} = \frac{dQ}{dz};\quad \mu\frac{d\beta}{dt} = -\frac{dP}{dz};\quad \mu\frac{d\gamma}{dt} = 0,\\ K\frac{dP}{dt} = -\frac{d\beta}{dz};\quad K\frac{dQ}{dt} = \frac{d\alpha}{dz};\quad K\frac{dR}{dt} = 0.\end{aligned}\right\}\text{.........(21).}$$

It follows that there is no component of either the electric or the magnetic force normal to the plane of the wave, and that therefore the whole of the disturbance is in that plane. If the electric disturbance is in one direction only, so that *e.g.* $Q = 0$, it follows that $\alpha = 0$, or that the magnetic disturbance is also rectilinear, and at *right angles to the electric disturbance.* We have therefore for the simplest case of a plane wave, two vectors representing the electric and magnetic forces respectively, and these vectors are at right angles to each other and to the direction of propagation.

More generally let the components P and Q of a plane wave-front parallel to xy be

$$P = \phi(z - vt);\quad Q = \psi(z - vt),$$

so that
$$v\frac{dP}{dz} = -\frac{dP}{dt};\quad v\frac{dQ}{dz} = -\frac{dQ}{dt};$$

or making use of (21)

$$\mu v\frac{d\beta}{dt} = \frac{dP}{dt};\quad \mu v\frac{d\alpha}{dt} = -\frac{dQ}{dt}\text{.....................(22).}$$

Hence:
$$\mu v\beta = P;\quad \mu v\alpha = -Q,$$

and
$$\alpha P + \beta Q = 0.$$

This shows that also in this more general case the electric and magnetic forces are at right angles to each other.

141. Double Refraction. In crystalline substances the specific inductive capacity of a plate may depend on the direction in which the plate is cut, relative to the axes of the crystals. The currents which are generated in such substances by a variation of electric force are not necessarily in the direction of the force, but if P, Q, R be the electric forces resolved in three directions at right angles to each other, and if the current in any one direction be a linear function of P, Q, R, then it may be proved that there are always three directions at right angles to each other such that the current is in the direction of the force. If we choose these directions for the coordinate axes, we may write

$$u = \frac{1}{4\pi} K_1 \frac{dP}{dt}, \qquad v = \frac{1}{4\pi} K_2 \frac{dQ}{dt}, \qquad w = \frac{1}{4\pi} K_3 \frac{dR}{dt} \quad \text{...(23)},$$

where K_1, K_2, K_3, are the three principal dielectric constants.

These equations replace (16). The elimination of α between (13) and (17) now leads to

$$\left.\begin{aligned} K_1 \mu \frac{d^2P}{dt^2} &= \nabla^2 P - \frac{d}{dx}\left(\frac{dP}{dx} + \frac{dQ}{dy} + \frac{dR}{dz}\right) \\ K_2 \mu \frac{d^2Q}{dt^2} &= \nabla^2 Q - \frac{d}{dy}\left(\frac{dP}{dx} + \frac{dQ}{dy} + \frac{dR}{dz}\right) \\ K_3 \mu \frac{d^2R}{dt^2} &= \nabla^2 R - \frac{d}{dz}\left(\frac{dP}{dx} + \frac{dQ}{dy} + \frac{dR}{dz}\right) \end{aligned}\right\} \quad \text{............(24).}$$

If $K_1 = K_2 = K_3$, $\qquad \dfrac{dP}{dx} + \dfrac{dQ}{dy} + \dfrac{dR}{dz} = 0,$

and we are brought back to the equations which have already been deduced for isotropic media. We proceed to investigate under what conditions plane waves are propagated in a medium to which equations (24) apply. If l, m, n, are the direction cosines of the normal of the plane wave, and V the velocity of propagation, all variable quantities must be expressible as functions of $lx + my + nz - Vt$.

We may therefore in the case of a rectilinear disturbance write for P, Q, R,

$$P_0 f(lx + my + nz - Vt), \quad Q_0 f(lx + my + nz - Vt),$$
$$R_0 f(lx + my + nz - Vt) \quad \text{......................(25)},$$

where P_0, Q_0, R_0 are constants defining the direction of the electric disturbance, the cosines of the angles which the direction of the electric force forms with the coordinate axes being as $P_0 : Q_0 : R_0$.

By substitution, equations (24) become, if we write

$$v_1 = 1/\sqrt{K_1 \mu}, \quad v_2 = 1/\sqrt{K_2 \mu}, \quad v_3 = 1/\sqrt{K_3 \mu}$$

and
$$S = lP + mQ + nR;$$

$$\left.\begin{aligned} P &= \frac{lSv_1^2}{v_1^2 - V^2} \\ Q &= \frac{mSv_2^2}{v_2^2 - V^2} \\ R &= \frac{nSv_3^2}{v_3^2 - V^2} \end{aligned}\right\} \qquad \text{..........................(26).}$$

Multiplying the first of these equations by l, the second by m, and the third by n, and adding, we obtain the characteristic equation for V,

$$\frac{v_1^2 l^2}{v_1^2 - V^2} + \frac{v_2^2 m^2}{v_2^2 - V^2} + \frac{v_3^2 n^2}{v_3^2 - V^2} = 1,$$

or subtracting
$$l^2 + m^2 + n^2 = 1,$$

$$\frac{l^2}{v_1^2 - V^2} + \frac{m^2}{v_2^2 - V^2} + \frac{n^2}{v_3^2 - V^2} = 0 \quad \text{..............(27).}$$

This is an equation identical with (4) Chapter VIII., and shows that the electromagnetic wave theory leads to the correct construction for the propagation of plane waves.

From (26) we also derive

$$\frac{lP}{v_1^2} + \frac{mQ}{v_2^2} + \frac{nR}{v_3^2} = \left(\frac{l^2}{v_1^2 - V^2} + \frac{m^2}{v_2^2 - V^2} + \frac{n^2}{v_3^2 - V^2}\right) S = 0.$$

As P/v_1^2, Q/v_2^2, R/v_3^2, are proportional to the components of electric current, we conclude that the electric current is in the plane of the wave-front.

The substitution of (25) into (13) leads to

$$V\mu\alpha = Rm - Qn,$$
$$V\mu\beta = Pn - Rl,$$
$$V\mu\gamma = Ql - Pm;$$

from which it follows that

$$l\alpha + m\beta + n\gamma = 0,$$

and
$$P\alpha + Q\beta + R\gamma = 0.$$

Hence the magnetic force is in the plane of the wave, and the electric force is at right angles to the magnetic force, though not in general, as will presently appear, in the plane of the wave.

In Art. 87 it was found that if an ellipsoid

$$v_1^2 x^2 + v_2^2 y^2 + v_3^2 z^2 = 1 \quad \text{..................... (28)}$$

be constructed, the reciprocals of the two principal axes of any plane section measure the two velocities of plane waves which are parallel to the section, and it was proved that this construction leads to equation (27). This equation has been shown to lead to Fresnel's wave-surface which is therefore now established as a consequence of the electromagnetic theory. Propositions with respect to wave or ray velocities which are proved in the same chapter may all be interpreted in terms of the electromagnetic theory if we take the components of the electric current u, v, w to correspond to the displacements in the older theory.

If an electric disturbance is propagated as a plane wave, and a normal be drawn to the ellipsoid (28) at the end of the vector having u, v, w as components, the direction cosines of this normal are proportional to v_1^2u, v_2^2v, v_3^2w and are therefore by (22) coincident with the direction cosines of the vector representing the electric force. This electric force is therefore not in the plane of the wave but lies in a plane which contains the wave normal and the electric current. It has been shown in Art. 89 that this plane also contains the ray.

142. Problem of refraction and reflexion. A good test of the adequacy of any theory of light is found in its capability of dealing with the problem of reflexion and refraction. Reflexion takes place when a wave falls on a surface at which the properties of the medium are suddenly changed. If the transition is gradual, there is no reflexion. A ray of light *e.g.* enters our atmosphere from outside and gradually passes into denser and denser layers of air. Though its path becomes curved by refraction, there is no reflexion, and neglecting absorption, the intensity of the ray remains unaltered. The fact that a surface of glass or water partially reflects a ray of light shows that the transition between the media of different refractive indices must take place within a distance not much greater than a wave-length.

Before entering into the relative merits of different theories with regard to the problem of reflexion, we may deduce some general results which are independent of any theory. We consider a plane wave-front having its normal in the plane xy. Its displacements, in whatever direction they are, must be capable of expression in the form $f(ax + by - ct)$, for

$$ax + by - ct = \text{constant}$$

expresses a plane parallel to the axis of z. If θ be the angle between the wave normal and the axis of x, and v the velocity of wave propagation, we have

$$\cos\theta = \frac{a}{\sqrt{a^2+b^2}}, \quad \sin\theta = \frac{b}{\sqrt{a^2+b^2}}, \quad \cot\theta = \frac{a}{b},$$

and

$$v = \frac{c}{\sqrt{a^2+b^2}}.$$

The line of intersection of the wave with the plane $x=0$ is

$$by - ct = \text{constant}$$

and travels forward therefore parallel to itself with a uniform speed of c/b.

If the plane $x=0$ is a surface at which refraction takes place, the displacement in the refracted wave may be expressed as

$$F(a_1x + b_1y - c_1t).$$

If the displacements are periodic, the periods must be the same in the refracted and incident beam, hence $c_1 = c$. Also the lines of intersection of the refracted and incident waves with the surface $x=0$ must always be coincident. Hence this line must travel forward with the same velocity in both waves. This proves that $b_1 = b$. The velocity of the refracted wave is

$$v_1 = \frac{c_1}{\sqrt{a_1^2 + b_1^2}} = \frac{c}{\sqrt{a_1^2 + b^2}}.$$

Hence calling the angle of refraction θ_1

$$\frac{v}{v_1} = \frac{\sqrt{a_1^2 + b^2}}{\sqrt{a^2 + b^2}} = \frac{\sin\theta}{\sin\theta_1}.$$

This proves the law of refraction. The displacements in the reflected wave will be of the form

$$F(a'x + b'y - c't).$$

The previous reasoning shows that $c' = c$, $b' = b$. Also the velocity of wave propagation must now be identical with that of the incident wave. Hence $a' = \pm a$. We must choose the lower sign, as otherwise the wave would simply return in the original direction. The numerical equality of a' and a proves the law of reflexion.

143. Reflexion in the Electromagnetic Theory. The problem of reflexion is comparatively simple if treated according to the electromagnetic theory, and we shall therefore begin with it. In the electrostatic or electromagnetic field the electric and magnetic forces have to satisfy certain conditions at the surface of separation of two media having different properties. These are in treatises on electricity proved to be the following: (1) The tangential components of *electric force* are the same on both sides of the surface. (2) The normal components of *electric displacement* are continuous. (3) The tangential components of *magnetic force* are continuous, and (4) the normal components of *magnetic induction* are continuous. Taking the surface $x=0$ to be the surface of separation, we may put with the previous notation these so-called surface conditions into the form:

$$K\frac{dP}{dt} = K_1\frac{dP_1}{dt}, \quad Q = Q_1, \quad R = R_1 \quad \ldots\ldots\ldots\ldots(29),$$

$$\mu\alpha = \mu_1\alpha_1, \qquad \beta = \beta_1, \qquad \gamma = \gamma_1 \ldots\ldots\ldots\ldots\ldots(30).$$

The right-hand sides of all equations apply to the second medium which we shall refer to as the *lower* medium, taking the axis x positive downwards.

These six equations are not all independent. The continuity of Q and R involves the continuity of their variations in any tangential direction and hence the first equation (13) shows that the continuity of normal magnetic induction is secured. Similarly the continuity of the tangential components of magnetic force leads to the continuity of the normal electric current as shown from the first of equations (18). We may therefore omit the first of equations (29) and (30) as being contained in the others; nevertheless it is often convenient to introduce them.

If the wave-front be parallel to the axis of z

$$dP/dz = 0.$$

Also writing without appreciable error $\mu_1 = \mu$ for all transparent media, the continuity of β is satisfied according to (13) if dR/dx is continuous. We may therefore replace the surface conditions by the following five:

$$\left.\begin{array}{c} KP = K_1P_1, \quad Q = Q_1, \quad R = R_1 \\ \dfrac{dR}{dx} = \dfrac{dR_1}{dx}, \quad \dfrac{dQ}{dx} - \dfrac{dP}{dy} = \dfrac{dQ_1}{dx} - \dfrac{dP_1}{dy} \end{array}\right\} \text{............(31).}$$

We now take the incident beam to be plane polarized and first treat the case that the electric force is at right angles to the plane of incidence which we take to be the plane of xy. Therefore $P = Q = 0$, and the surface conditions reduce to

$$\left.\begin{array}{c} R = R_1 \\ \dfrac{dR}{dx} = \dfrac{dR_1}{dx} \end{array}\right\} \text{..........................(32).}$$

For the electric force in the incident wave we may write $e^{i(ax+by-ct)}$ and for that of the reflected wave $re^{i(-ax+by-ct)}$, where the real parts only are ultimately retained. A change of phase will be indicated by a complex value of r, or by a negative sign if the change is equal to π. If s is the amplitude of the transmitted beam, we have therefore

$$\left.\begin{array}{l} R = e^{i(ax+by-ct)} + re^{i(-ax+by-ct)} \\ R_1 = se^{i(a_1x+by-ct)} \end{array}\right\} \text{..................(33),}$$

in the upper and lower media respectively.

The surface conditions give at once for $x = 0$

$$1 + r = s,$$

$$a(1 - r) = a_1 s,$$

$$r = \frac{a - a_1}{a + a_1}$$

$$= \frac{\cot\theta - \cot\theta_1}{\cot\theta + \cot\theta_1}$$

$$= \frac{\sin(\theta_1 - \theta)}{\sin(\theta_1 + \theta)} \text{.......................(34).}$$

The square of this expression correctly represents the observed intensity of the reflected beam, if the incident beam is polarized in the plane of incidence. We conclude that the electric force is at right angles to the plane of polarization, a result in accordance with the conclusion arrived at in the study of double refraction.

If we take the incident beam to be polarized at right angles to the plane of incidence, R, α, β vanish, and the surface conditions become

$$\left.\begin{aligned} &KP = K_1P_1, \quad Q = Q_1 \\ &\frac{dQ}{dx} - \frac{dP}{dy} = \frac{dQ_1}{dx} - \frac{dP_1}{dy} \end{aligned}\right\} \qquad (35).$$

The last equation secures the continuity of γ. But the form of our assumed disturbance shows that $d\gamma/dy = ib\gamma$ and hence if γ is continuous so is also $d\gamma/dy$ and vice versa. Also according to (18) $K\,dP/dt = d\gamma/dy$, when $\beta = 0$: the first and last surface conditions are therefore identical and we may disregard the latter.

If WF (Fig. 169) be the incident wave-front, the displacement is now in the plane of the paper and parallel to the wave-front. Let the direction indicated by the arrow be that in which the displacements are taken to be positive. $W'F'$ represents the reflected wave-front, and we may again arbitrarily fix that direction for which we shall take the displacements to be positive.

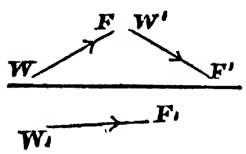


Fig. 169.

It is obvious that for normal incidence there is no distinction between this case and the one already considered when the displacement is at right angles to the plane of incidence. It is therefore convenient to take that direction as positive which agrees with that of the incident wave when the incidence is normal. The arrow indicates the direction. Similarly for the transmitted wave W_1F_1. Taking the amplitude of the incident beam again to be unit amplitude, and resolving along OX and OY, we may put in the upper medium

$$P = -\sin\theta\, e^{i(ax+by-ct)} + r\sin\theta\, e^{i(-ax+by-ct)},$$

$$Q = \cos\theta\, e^{i(ax+by-ct)} + r\cos\theta\, e^{i(-ax+by-ct)},$$

and in the lower medium

$$P_1 = -s\sin\theta_1 e^{(a_1x+by-ct)},$$

$$Q_1 = s\cos\theta_1 e^{(a_1x+by-ct)}.$$

The condition $KP = K_1P_1$ for $x = 0$ gives

$$\frac{(1-r)}{v^2}\sin\theta = \frac{s\sin\theta_1}{v_1^{\,2}},$$

or

$$(1-r)\sin\theta_1 = s\sin\theta,$$

and the condition $Q = Q_1$ gives

$$(1+r)\cos\theta = s\cos\theta_1.$$

These are the only conditions that need be satisfied.

Eliminating s we obtain

$$(1-r)\sin\theta_1\cos\theta_1=(1+r)\sin\theta\cos\theta,$$

or

$$r=\frac{\sin\theta_1\cos\theta_1-\sin\theta\cos\theta}{\sin\theta_1\cos\theta_1+\sin\theta\cos\theta}$$

$$=\frac{\sin 2\theta_1-\sin 2\theta}{\sin 2\theta_1+\sin 2\theta}$$

$$=\frac{\tan(\theta_1-\theta)}{\tan(\theta_1+\theta)}\quad\text{..............................}(36).$$

This again is a formula agreeing with observation, at any rate as a first approximation. The application of the equations (34) and (36) to the cases of oblique polarization or unpolarized light has already been discussed in Art. 27 as well as the observed departures from (36).

It has often been suggested that the experimental deviations from the tangent law may be due to the fact that the transition between the two media is not sudden but takes place within a layer comparable in thickness with the length of a wave. L. Lorenz* first investigated the question and showed that a thickness of from the tenth to the hundredth part of a wave-length is sufficient to cause the observed effect. Drude†, treating the same subject from the standpoint of the electromagnetic theory, has arrived at similar results, a thickness of the transition layer of ·0175 λ being found to be sufficient in the case of flint-glass to account for the elliptic polarization observed near the polarizing angle‡.

144. Reflexion in the elastic solid theory. In elastic solids the conditions at the boundary are obtained by the consideration that as a tearing of the medium can only take place under application of forces which exceed the limits of elasticity, the displacements on both sides of the boundary must be the same, while the medium is performing oscillations under the conditions of perfect elasticity.

A second condition is imposed by the third law of motion. *The stresses must be continuous.* The continuity of stress together with that of displacement satisfies also the requirements of the law of conservation of energy, as the work done across any surface is the product of stress and rate of change of displacement.

The components of displacement which we had previously called α, β, γ, shall, in order to distinguish them from the magnetic forces for which we have introduced the same letters, now be designated by ξ, η, ζ.

* *Pogg. Ann.* CXI. p. 460 (1860) and CXIV. p. 238 (1861).

† *Lehrbuch der Optik*, p. 266.

‡ The most complete investigation of the subject has been given by Maclaurin, *Proc. Roy. Soc.* LXXVI. p. 49 (1905).

The continuity of displacement introduces the conditions

$$\xi = \xi_1, \quad \eta = \eta_1, \quad \zeta = \zeta_1,$$

where the right-hand sides refer to the lower medium.

The stresses on a surface normal to the axis of x are, by Art. 134,

$$P = A\frac{d\xi}{dx} + B\left(\frac{d\eta}{dy} + \frac{d\zeta}{dz}\right);$$

$$T = n\left(\frac{d\zeta}{dx} + \frac{d\xi}{dz}\right); \quad U = n\left(\frac{d\xi}{dy} + \frac{d\eta}{dx}\right),$$

where $\qquad A = k + \frac{4}{3}n; \quad B = k - \frac{2}{3}n.$

Writing $m = k + \frac{1}{3}n$, we obtain for the conditions of continuity of stress

$$\left.\begin{aligned} &(m+n)\frac{d\xi}{dx} + (m-n)\left(\frac{d\eta}{dy} + \frac{d\zeta}{dz}\right) \\ &\qquad = (m_1+n_1)\frac{d\xi_1}{dx} + (m_1-n_1)\left(\frac{d\eta_1}{dy} + \frac{d\zeta_1}{dz}\right) \\ &n\left(\frac{d\zeta}{dx} + \frac{d\xi}{dz}\right) = n_1\left(\frac{d\zeta_1}{dx} + \frac{d\xi_1}{dz}\right) \\ &n\left(\frac{d\xi}{dy} + \frac{d\eta}{dx}\right) = n_1\left(\frac{d\xi_1}{dy} + \frac{d\eta_1}{dx}\right) \end{aligned}\right\} \quad \ldots\ldots\ldots\ldots\ldots(37),$$

where m_1, n_1, define the elastic properties of the second medium.

Let the plane of xy be the plane of incidence, and the vibrations of a plane wave be at right angles to that plane. All displacements vanish except ζ, and ζ is independent of z. Hence the equations of continuity reduce to

$$\zeta = \zeta_1; \quad n\frac{d\zeta}{dx} = n_1\frac{d\zeta_1}{dx}.$$

The equation of motion in the upper medium is, according to (9),

$$\frac{d^2\zeta}{dt^2} = \frac{n}{\rho}\left(\frac{d^2\zeta}{dx^2} + \frac{d^2\zeta}{dy^2}\right),$$

with a similar equation for the lower medium.

But $$\zeta = e^{i(ax+by-ct)} + re^{i(-ax+by-ct)},$$

$$\zeta_1 = se^{i(a_1x+by-ct)}.$$

For $x = 0$, the surface conditions give

$$1 + r = s,$$

$$na(1-r) = n_1a_1s,$$

and eliminating s,

$$r = \frac{na - n_1a_1}{na + n_1a_1} = \frac{n\cot\theta - n_1\cot\theta_1}{n\cot\theta + n_1\cot\theta_1} \ldots\ldots\ldots\ldots\ldots(38).$$

For the velocity of wave propagation in an elastic solid, we have $v^2 = n/\rho$. Different wave velocities in different media may either be due to differences in the rigidity or to differences in density. Hence we must distinguish the two cases.

Case I. $n = n_1$.

Equation (33) becomes

$$r = \frac{\cot\theta - \cot\theta_1}{\cot\theta + \cot\theta_1} = \frac{\sin(\theta_1 - \theta)}{\sin(\theta + \theta_1)}.$$

This agrees with the result obtained in the electromagnetic theory if the displacements are made to correspond to electric force.

Case II.

$$\rho = \rho_1; \quad \frac{n_1}{n} = \frac{v_1^2}{v^2} = \frac{\sin^2\theta_1}{\sin^2\theta}.$$

Equation (33) now gives

$$r = \frac{\sin^2\theta\tan\theta_1 - \sin^2\theta_1\tan\theta}{\sin^2\theta\tan\theta_1 + \sin^2\theta_1\tan\theta} = \frac{\sin 2\theta - \sin 2\theta_1}{\sin 2\theta + \sin 2\theta_1} = \frac{\tan(\theta - \theta_1)}{\tan(\theta + \theta_1)}.$$

This is the equation for the reflected light when the incident wave is polarized at right angles to the plane of incidence. Hence if different media differ by their rigidities, the reflexion of light vibrating at right angles to the plane of incidence can only be accounted for by supposing that the plane of polarization contains the vibration.

To work out completely the more complicated case that the vibration lies in the plane of incidence, we must transform the equations of motion.

In equations (9) alter the notation, put $\zeta = 0$, and let ξ and η be independent of z, this being the condition that the wave normal lies in the plane of xy. The equations then become

$$\rho\frac{d^2\xi}{dt^2} = m\frac{d}{dx}\left(\frac{d\xi}{dx} + \frac{d\eta}{dy}\right) + n\left(\frac{d^2\xi}{dx^2} + \frac{d^2\xi}{dy^2}\right),$$

or

$$\left.\begin{aligned} \rho\frac{d^2\xi}{dt^2} &= (m+n)\frac{d}{dx}\left(\frac{d\xi}{dx} + \frac{d\eta}{dy}\right) + n\frac{d}{dy}\left(\frac{d\xi}{dy} - \frac{d\eta}{dx}\right) \\ \text{Similarly}\quad \rho\frac{d^2\eta}{dt^2} &= (m+n)\frac{d}{dy}\left(\frac{d\xi}{dx} + \frac{d\eta}{dy}\right) - n\frac{d}{dx}\left(\frac{d\xi}{dy} - \frac{d\eta}{dx}\right) \end{aligned}\right\} \quad \ldots(39).$$

Introducing two new functions such that

$$\xi = \frac{d\phi}{dx} + \frac{d\psi}{dy}; \quad \eta = \frac{d\phi}{dy} - \frac{d\psi}{dx} \quad \ldots\ldots\ldots\ldots\ldots\ldots(40),$$

we find that (39) may be satisfied by

$$\left.\begin{aligned}\rho\frac{d^2\phi}{dt^2}&=(m+n)\left(\frac{d^2\phi}{dx^2}+\frac{d^2\phi}{dy^2}\right)\\ \text{and}\qquad \rho\frac{d^2\psi}{dt^2}&=n\left(\frac{d^2\psi}{dx^2}+\frac{d^2\psi}{dy^2}\right)\end{aligned}\right\}\quad\text{..............(41).}$$

From (35) it appears that the displacements ξ and η, due to changes of ϕ, are at right angles to the surface $\phi=$ constant, while the displacements due to changes in ψ lie *in* the surface $\psi=$ constant. If we adopt the same form of solution for ϕ and ψ, it is the latter function which gives the motion which we require for the propagation of light in which the displacements are in the wave-front. We put therefore for the incident wave $\psi=e^{i(ax+by-ct)}$, and assume for the form of solution generally,

In the upper medium:

$$\left.\begin{aligned}\psi&=e^{i(ax+by-ct)}+re^{i(-ax+by-ct)}\\ \phi&=pe^{i(a'x+by-ct)}\end{aligned}\right\}\quad\text{.................(42).}$$

In the lower medium:

$$\left.\begin{aligned}\psi_1&=se^{i(a_1x+by-ct)}\\ \phi_1&=qe^{i(a_1'x+by-ct)}\end{aligned}\right\}\quad\text{..............................(43).}$$

Substituting these values in equations (41) we obtain

$$\begin{aligned}c^2=\frac{n}{\rho}(a^2+b^2)=\frac{n_1}{\rho_1}(a_1^2+b^2)&=\frac{(m+n)}{\rho}(a'^2+b^2)\\ &=\frac{m_1+n_1}{\rho_1}(a_1'^2+b^2)\quad\text{......(44).}\end{aligned}$$

From the first two equalities we obtain as before the law of refraction, but as m and m_1 are indefinitely great the last equalities give

$$a'^2+b^2=0\;;\quad a_1'^2+b^2=0\quad\text{.................(45).}$$

For the same reason

$$\frac{d^2\phi}{dx^2}+\frac{d^2\phi}{dy^2}=\frac{d^2\phi_1}{dx^2}+\frac{d^2\phi_1}{dy^2}=0\text{..................(46).}$$

This shows that the motion due to ϕ is that of an incompressible liquid. As ϕ represents the velocity potential, the motion is irrotational. Also by substitution of (45) into (42) and (43) retaining only the real parts:

$$\phi=pe^{-bx}\cos(by-ct),$$
$$\phi_1=qe^{-bx}\cos(by-ct).$$

The displacements in so far as they are due to ϕ are

$$\frac{d\phi}{dx}=-bpe^{-bx}\cos(by-ct)\qquad\text{(upper medium),}$$

$$\frac{d\phi_1}{dx}=-bqe^{-bx}\cos(by-ct)\qquad\text{(lower medium),}$$

and $$\frac{d\phi}{dy} = -bpe^{-bx}\sin(by - ct) \qquad \text{(upper medium)},$$

$$\frac{d\phi_1}{dy} = -bqe^{-bx}\sin(by - ct) \qquad \text{(lower medium)}.$$

The motion vanishes for normal incidence as $b = \frac{2\pi}{\lambda}\sin\theta$. Unless θ is small, the exponential factor shows that the motion quickly diminishes with the distance from the refracting surface.

The surface conditions are

$$\xi = \xi_1\,; \quad \eta = \eta_1\,;$$

$$(m + n)\frac{d\xi}{dx} + (m - n)\frac{d\eta}{dy} = (m_1 + n_1)\frac{d\xi_1}{dx} + (m_1 - n_1)\frac{d\eta_1}{dy},$$

or in terms of ψ and ϕ

$$(m + n)\frac{d^2\phi}{dx^2} + (m - n)\frac{d^2\phi}{dy^2} + 2n\frac{d^2\psi}{dx\,dy} = \text{similar expressions},$$

or introducing (46)

$$n\left\{2\frac{d^2\psi}{dx\,dy} + \frac{d^2\phi}{dx^2} - \frac{d^2\phi}{dy^2}\right\} = \text{similar expressions}.$$

The quantities p, q, r, s may now be obtained by substituting ϕ, ψ, ϕ_1, ψ_1 from equations (42) and (43). Green*, to whom the above investigation is due, assumes $n = n_1$, and Rayleigh† has put the solution for r in that case into the form

$$r^2 = \frac{\cot^2(\theta + \theta') + M^2}{\cot^2(\theta - \theta') + M^2},$$

where $M = \frac{\mu^2 - 1}{\mu^2 + 1}$ and μ is the refractive index.

If two media do not differ much in optical properties, so that the refractive index is nearly equal to one, we obtain for the ratio of amplitudes the expression

$$\frac{\tan(\theta - \theta')}{\tan(\theta + \theta')},$$

as required by experiment when the vibration takes place at right angles to the plane of incidence.

As has been pointed out above, the tangent formula is only approximately correct, but the deviations are not so great as those which Green's formula would lead us to expect, and are sometimes in the other direction.

The alternative according to which differences in optical properties are due to differences in elasticity, leads to results which can in no way be reconciled with observed facts. If we place ourselves on the standpoint of the elastic solid theory, we are therefore compelled to conclude

* *Collected Works*, p. 245. † *Collected Works*, vol. I. p. 129.

that the rigidity of the æther is the same in all media. Even then we arrive at an unsatisfactory result so far as light polarized at right angles to the plane of incidence is concerned.

145. Lord Kelvin's theory of contractile æther. According to the most general equations of the motion of an elastic substance (Art. 135), a disturbance spreads in the form of two waves, the condensational longitudinal wave propagated with a velocity $\sqrt{(k+\frac{4}{3}n)/\rho}$ and the transverse distortional wave propagated with a velocity $\sqrt{n/\rho}$. The phenomena of light leave no room for a longitudinal wave propagated with finite velocity. The theories so far considered avoid the difficulty by taking the elastic body as incompressible, when the coefficient k becomes infinitely large, and the longitudinal disturbance is propagated with infinite velocity.

This elastic solid theory of the æther which has been discussed in the preceding articles, does not, as appeared, consistently lead to facts which are in agreement with observation: it fails to account for the laws of double refraction and for the observed amplitude of light reflected by transparent bodies. That theory was therefore considered dead, until Lord Kelvin* resuscitated it in a different form by showing how, dropping the hypothesis of "solidity," an elastic theory of the æther may still be a possible one.

The characteristic distinction of the new theory lies in the bold assumption that the velocity of the longitudinal wave, instead of being infinitely large, is infinitely small. This requires that $k+\frac{4}{3}n$ shall be zero, so that k is negative. A medium in which there is a negative resistance to compression would at first sight appear to be essentially unstable, but Lord Kelvin shows that the instability cannot come into play, if the æther is rigidly attached to a bounding surface. So long as there is a finite propagational velocity for each of the two kinds of wave motion, no disturbance set up in the medium can lead to instability. Putting therefore the constant A of Article 134 equal to zero, and taking the rigidity to be equal in all media, Lord Kelvin has shown that the theory leads to Fresnel's tangent formula for the amplitude of light polarized in a plane perpendicular to the plane of incidence. Glazebrook† then showed that the consideration of double refraction leads to Fresnel's wave surface, while J. Willard Gibbs‡ pointed out that the new form of elastic æther theory must always lead to the same equation as the electromagnetic theory, provided we replace the symbol which denotes "displacement" in one theory by that which denotes "force" in the other and vice versa.

* *Phil. Mag.* XXVI. p. 414. 1888.
† *Phil. Mag.* XXVI. p. 521. 1888.
‡ *Phil. Mag.* XXVII. p. 238. 1889.

If in equations (9) we write $k = -\frac{4}{3}n$, and allow different values of ρ according as the displacements are in the direction x, y, or z, the equations become with our present notation

$$\left.\begin{aligned}\frac{\rho_1}{n}\frac{d^2\xi}{dt^2} &= \nabla^2\xi - \frac{d}{dx}\left(\frac{d\xi}{dx} + \frac{d\eta}{dy} + \frac{d\zeta}{dz}\right)\\ \frac{\rho_2}{n}\frac{d^2\eta}{dt^2} &= \nabla^2\eta - \frac{d}{dy}\left(\frac{d\xi}{dx} + \frac{d\eta}{dy} + \frac{d\zeta}{dz}\right)\\ \frac{\rho_3}{n}\frac{d^2\zeta}{dt^2} &= \nabla^2\zeta - \frac{d}{dz}\left(\frac{d\xi}{dx} + \frac{d\eta}{dy} + \frac{d\zeta}{dz}\right)\end{aligned}\right\} \quad \text{............(47).}$$

These equations are identical with (24) provided that we replace P, Q, R in the latter by ξ, η, ζ, and μ, K_1, K_2, K_3 by $1/n$, ρ_1, ρ_2, ρ_3, respectively. As regards surface conditions, we must now remember that the resistance to compression being negative, there may be infinite compression or dilatation at any point or surface at which a condensational wave tends to start. The surface at which reflexion takes place gives rise according to the preceding article to condensational waves, hence disregarding this wave which can only be propagated with zero velocity, the conditions which hold in the general elastic theory, in which the condensational wave is considered, are not necessarily satisfied. They must be replaced by others, which, as J. W. Gibbs has shown, may be obtained directly from the equations of motion.

Introduce new quantities defined by

$$\xi' = \frac{d\zeta}{dy} - \frac{d\eta}{dz}, \qquad \eta' = \frac{d\xi}{dz} - \frac{d\zeta}{dx}, \qquad \zeta' = \frac{d\eta}{dx} - \frac{d\xi}{dy},$$

$$\xi'' = \frac{d\zeta'}{dy} - \frac{d\eta'}{dz}, \qquad \eta'' = \frac{d\xi'}{dz} - \frac{d\zeta'}{dx}, \qquad \zeta'' = \frac{d\eta'}{dx} - \frac{d\xi'}{dy}.$$

Performing the differentiation so as to obtain ξ'', η'', ζ'' in terms of ξ, η, ζ, we arrive at the expressions which stand on the right-hand side of (47) with the sign reversed. We take the boundary to be normal to the axis of x. If η' and ζ' were discontinuous at such a boundary, $\frac{d\eta'}{dx}$ and $\frac{d\zeta'}{dx}$ would be infinitely large, and this would make η'' and ζ'', and consequently also $\frac{d^2\eta}{dt^2}$, $\frac{d^2\zeta}{dt^2}$ infinitely large, which is obviously not admissible. Hence we conclude that η' and ζ' are continuous. A similar reasoning shews that as η' and ζ' must remain finite, η and ζ must be continuous. The conditions of continuity are therefore

$$\eta = \eta_1; \ \zeta = \zeta_1,$$

$$\frac{d\xi}{dz} - \frac{d\zeta}{dx} = \frac{d\xi_1}{dz} - \frac{d\zeta_1}{dx}; \ \frac{d\eta}{dx} - \frac{d\xi}{dy} = \frac{d\eta_1}{dx} - \frac{d\xi_1}{dy}.$$

But these are exactly the conditions which we have seen must hold in the electromagnetic theory if ξ, η, ζ are replaced by P, Q, R. The analogy is made complete if it is noticed that the continuity of ξ' and ξ'' follows from the above equations, and that according to the first of equations (47) if $v_1 = \sqrt{n}/\rho$,

$$\frac{1}{v_1^2}\frac{d^2\xi}{dt^2} = -\xi''.$$

The continuity of ξ'' is seen to involve the continuity of the normal force and this corresponds to the continuity of electric displacement in the electromagnetic theory.

The analogy has therefore been proved both for the equations regulating the motion of the medium and for the surface conditions.

If this analogy is kept in view, all the results which have been found to hold in the electromagnetic theory may be translated at once into consequences of the contractile æther theory. Thus in the theory of double refraction, the displacements of the latter theory are not in the wave surface, but are normal to the ray as has been shown for the electric forces in Art. 136. Similarly adopting the contractile æther theory we may conclude that when plane waves are propagated through a doubly refracting medium the elastic force and not the displacement is in the plane of the wave. I have given a statement of this theory on account of its mathematical interest, but it was ultimately abandoned by its author*.

146. Historical. AUGUSTIN LOUIS CAUCHY, born August 21st, 1789, in Paris, died May 23rd, 1857, at Sceaux, near Paris, was one of the large number of celebrated French mathematicians who, during the end of the 18th and the beginning of the 19th century, made the first serious advance in Mathematical Physics since Newton's time. Cauchy's contribution to the theory of light consisted in initiating the endeavour to deduce the differential equations for the motion of light from a theory of elasticity. This theory was based on definite assumptions of the actions between the ultimate particles of matter. The luminiferous æther like other matter was supposed to be made up of distinct centres of force acting upon each other according to some law depending on the distance. Cauchy explained the phenomena of dispersion by supposing that in the media in which dispersion takes place, the distance between the ultimate particles is no longer small compared with the wave-length. He thus arrived at a formula which for a long time was considered to represent satisfactorily the connexion between wave-length and refractive index (Art. 153). Cauchy also showed that metallic reflexion may be accounted for by a high

* *Baltimore Lectures*, p. 214.

coefficient of absorption: by interpreting Fresnel's sine and tangent formula, in the case where the index of refraction is imaginary, he obtained the equations for the elliptic polarization of light reflected from metallic surfaces, which are still adopted as correctly representing the facts.

George Green was born at Sneinton, near Nottingham, in 1793, and only entered the University of Cambridge at the age of 40. Having graduated in 1837 as fourth wrangler, he was elected to a fellowship in Gonville and Caius College in 1839, and died in 1841. The following paragraph which stands at the head of his celebrated Memoir on the Reflexion and Refraction of Light will show the ideas which guided him in his work.

"M. Cauchy seems to have been the first who saw fully the utility of applying to the Theory of Light those formulæ which represent the motions of a system of molecules acting on each other by mutually attractive and repulsive forces; supposing always that in the mutual action of any two particles, the particles may be regarded as points animated by forces directed along the right line which joins them. This last supposition, if applied to those compound particles, at least, which are separable by mechanical division, seems rather restrictive; as many phenomena, those of crystallisation for instance, seem to indicate certain polarities in these particles. If, however, this were not the case, we are so perfectly ignorant of the mode of action of the elements of the luminiferous ether on each other, that it would seem a safer method to take some general physical principle as the basis of our reasoning, rather than assume certain modes of action, which, after all, may be widely different from the mechanism employed by nature; more especially if this principle include in itself, as a particular case, those before used by M. Cauchy and others, and also lead to a much more simple process of calculation. The principle selected as the basis of the reasoning contained in the following paper is this: In whatever way the elements of any material system may act upon each other, if all the internal forces exerted be multiplied by the elements of their respective directions, the total sum for any assigned portion of the mass will always be the exact differential of some function. But, this function being known, we can immediately apply the general method given in the *Mécanique Analytique*, and which appears to be more especially applicable to problems that relate to the motions of systems composed of an immense number of particles mutually acting upon each other. One of the advantages of this method, of great importance, is, that we are necessarily led by the mere process of the calculation, and with little care on our part, to all the equations and conditions which are requisite and sufficient for the complete solution of any problem to which it may be applied."

The function introduced above by Green we now call "Potential Energy," and a particular interest attaches to the whole paper, as it is the first instance of the application of the principle of Conservation of Energy to a great physical problem. Green shows that in the most general case, there may be twenty-one different coefficients defining the elastic properties of a medium, and that these reduce to two in the case of an isotropic or uncrystallized medium. The conditions which hold at the surface of two media are deduced, and for the first time strict dynamic principles were applied to the calculation of the amplitudes of the reflected and refracted light. Assuming the difference in the optical behaviour of different media to be differences of density, the Fresnel sine formula is obtained for light polarized in the plane of incidence, and it is shown that for light polarized at right angles to the plane of incidence, the tangent formula can only hold approximately. In a further paper "On the propagation of light in crystallized media," we meet the difficulties which have so long beset all attempts to account satisfactorily for Fresnel's wave surfaces, and though this paper will still be read with advantage, its interest at present is only historical.

George Gabriel Stokes, born August 13th, 1819, at Screen in Ireland, graduated as Senior Wrangler in 1841, and was elected to the Lucasian Chair of Mathematics in Cambridge in 1849. He died on February 1st, 1903. His celebrated Memoir "On the Dynamical Theory of Diffraction" contains the complete solution of the problem of the propagation of waves through an elastic medium. The question is treated in so masterly a manner that though published in the year 1849, the paper should still be carefully studied by every student of Optics. He published other important optical memoirs, of which the following may specially be quoted: "On the theory of certain bands seen in the spectrum" (1848); "On the formation of the Central Spot of Newton's Rings beyond the critical angle" (1848); "On Haidinger's Brushes" (1850); "Report on Double Refraction" (1862).

Many of his writings on the theory of sound and hydrodynamics have also optical applications. Stokes was the first to recognize the true nature of fluorescence, only isolated facts as to the luminescence of certain substances under the action of light having been previously known. He made a thorough experimental investigation which proved the possibility of a change in the refrangibility of light (*Phil. Trans.* 1852 and 1853).

James Clerk Maxwell, born June 13th, 1831, at Edinburgh, died at Cambridge November 5th, 1879, was the first occupant of the

Cavendish Professorship of Physics at Cambridge, which he held from the date of its foundation in 1871 to the time of his death. He was one of the most original minds who ever turned their attention to scientific enquiry. All mathematicians who, previous to Maxwell, had discussed the undulatory theory of Optics, started from the elastic solid theory of the luminiferous æther. That theory was able to give a satisfactory account of a great number of the phenomena of light and was considered to be securely established. The phenomena of electricity were treated as independent facts, though no doubt many physicists held that ultimately electric action would be explained by the stresses and strains of the same medium which transmitted light. No one had, however, suggested properties of the medium different from those of an ordinary elastic solid. Maxwell attacked the question with great originality from another point of view. Having asked himself the question, what the properties of a medium must be, in order that it should be capable of transmitting electric actions, he discovered that this electric medium was capable of transmitting transverse vibrations with the velocity of light. Maxwell also showed how Fresnel's wave-surface in double refracting media could be obtained by assuming that, in such media, there may be three dielectric constants, the polarization measured along three axes at right angles to each other being different. Of his other optical writings, his memoir "On the theory of Compound Colours and the relations of the Colours of the Spectrum" (*Phil. Trans.* 1860) deserves special mention.

CHAPTER XI.

DISPERSION AND ABSORPTION.

147. Wave-fronts with varying amplitudes. We have hitherto confined our attention to vibrations having the same amplitude along each wave-front. In other words, the surfaces of equal phase were coincident with the surfaces of equal amplitude. We shall now treat the question in a more general manner, starting from the differential equation of the wave propagation in an absorbing medium which, as we shall see, may be put into the form

$$G\frac{d^2R}{dt^2} + F\frac{dR}{dt} = \frac{d^2R}{dx^2} + \frac{d^2R}{dy^2} + \frac{d^2R}{dz^2} \quad \text{..............(1)},$$

where G and F are constants and R represents the displacement in the elastic solid theory or the electric force in the electromagnetic theory measured in the z direction. If the wave-front is plane and parallel to the axis of z, R is independent of z. If the disturbance is simply periodic, so that the time only occurs in the form of a periodic factor $pe^{-i\omega t}$, where p may be real or imaginary or complex, equation (1) is equivalent to

$$-(iF\omega + G\omega^2)R = \frac{d^2R}{dx^2} + \frac{d^2R}{dy^2},$$

or

$$\frac{d^2R}{dx^2} + \frac{d^2R}{dy^2} + \vartheta^2 R = 0 \quad \text{........................(2)},$$

where ϑ^2 depends on the constants of the medium and on the frequency; ϑ^2 is real when F is zero and only in that case. For a particular solution of (2) and therefore of (1), we have

$$R = R_0 e^{i(ax+by-\omega t)} \quad \text{........................(3)}.$$

The substitution of this value of R into (2) leads to the condition

$$a^2 + b^2 = \vartheta^2 \quad \text{..............................(4)},$$

showing that $a^2 + b^2$ must be independent of the *direction* in which the wave is propagated. If a and b are both real, there is no absorption and we may put

$$a = \frac{2\pi}{\lambda'}\cos\theta, \quad b = \frac{2\pi}{\lambda'}\sin\theta; \quad \omega = \frac{2\pi v}{\lambda'},$$

where λ' is the wave-length in the medium. The planes of equal phase are represented by

$$x \cos \theta + y \sin \theta = \text{constant}.$$

Let, in the more general case, a and b be complex and write therefore

$$a = \frac{2\pi}{\lambda'} \cos \theta + k_1 i,$$

$$b = \frac{2\pi}{\lambda'} \sin \theta + k_2 i.$$

Then (3) may be written, retaining the real parts,

$$R = R_0 e^{-(k_1 x + k_2 y)} \cos \frac{2\pi}{\lambda'} (x \cos \theta + y \sin \theta - vt) \quad \text{.........(5).}$$

The amplitude is the same over planes satisfying the equation

$$k_1 x + k_2 y = \text{constant},$$

but these planes do not necessarily coincide with the planes of equal phase. We have, however, still to satisfy the condition (4). It will be convenient here for the sake of obtaining symmetrical expressions to write

$$k_1 = \frac{2\pi}{\lambda} \kappa \cos \alpha,$$

$$k_2 = \frac{2\pi}{\lambda} \kappa \sin \alpha,$$

where λ is in vacuo the wave-length corresponding to a given ω, *i.e.* $\lambda = 2\pi V/\omega$ (V being the velocity of light in vacuo). We also write

$$\frac{V}{v} = \frac{\lambda}{\lambda'} = \nu \quad \text{..............................(6).}$$

Hence
$$a = \frac{2\pi}{\lambda} (\nu \cos \theta + i\kappa \cos \alpha),$$

$$b = \frac{2\pi}{\lambda} (\nu \sin \theta + i\kappa \sin \alpha),$$

$$a^2 + b^2 = \frac{4\pi^2}{\lambda^2} \{\nu^2 - \kappa^2 + 2i\nu\kappa \cos (\theta - \alpha)\} \quad \text{............ (7).}$$

In perfectly transparent media $a^2 + b^2$ is real, hence $\kappa = 0$ or $\cos(\theta - \alpha) = 0$. The first alternative leads to the case already discussed, of waves of equal amplitude. But the second alternative shows the possibility of waves of unequal amplitude being transmitted as plane waves, provided that the surfaces of equal amplitude are at right angles to the surfaces of equal phase. If we take the axis of x for the direction of propagation, (5) takes the form

$$R = R_0 e^{-\frac{2\pi}{\lambda}\kappa y} \cos \frac{2\pi\nu}{\lambda} (x - vt) \quad \text{..................(8).}$$

One important and somewhat unexpected result follows: *The velocity with which the wave-front proceeds depends on the variation of amplitude along it.* This is shown by (7) for

$$\vartheta^2 = a^2 + b^2 = \frac{4\pi^2}{\lambda^2}(\nu^2 - \kappa^2) \quad \text{.....................(9)},$$

$$\therefore \nu^2 = \frac{\lambda^2\vartheta^2}{4\pi^2} + \kappa^2 = \nu_1^2 + \kappa^2,$$

if ν_1 be the particular value which ν takes when κ is zero. It now follows from (6) that for v the velocity of the wave-front we have

$$v = \frac{V}{\sqrt{\nu_1^2 + \kappa^2}}.$$

The velocity with which a disturbance is transmitted is represented by the *ray* velocity V/ν_1 which must of course always be the same in the same medium. But our investigation is important as showing that even in vacuo the ray need not be at right angles to the wave-front, and that if this is the case the velocity of the wave-front is not the velocity with which a disturbance is transmitted. If ϕ be the angle between the ray and the wave normal

$$\cos\phi = \frac{\nu_1 v}{V} = \frac{\nu_1}{\sqrt{\nu_1^2 + \kappa^2}}.$$

In order to obtain an idea of the magnitude of the effect, assume ϕ to be equal to 1°; a short calculation then shows that the amplitude along the wave-front must be reduced in the ratio of $e:1$ along a distance equal to the 200th part of a millimetre. A more interesting case arises when κ and therefore κ_1 and κ_2 are imaginary, the amplitude of the wave represented by equation (5) is then a periodic function of x and y, and the rate of propagation becomes $V/(\nu^2 - \kappa^2)^{\frac{1}{2}}$. Such waves have been called "corrugated waves" by the late Lord Rayleigh*, who has made use of them in some of his investigations on the theory of gratings. When absorption takes place ϑ^2 is no longer real. Substituting for that quantity its value in terms of F and G, and combining (4) and (7), we find

$$\left.\begin{aligned} \nu^2 - \kappa^2 &= GV^2 \\ 2\nu\kappa\cos\rho &= \frac{\lambda}{2\pi}FV \end{aligned}\right\} \quad \text{.....................(10)},$$

where ρ represents the angle between the planes of equal phase and the planes of equal amplitude.

Both ν and κ now depend on the angle ρ, and I shall write ν_0, κ_0 for the values of ν, κ in the particular case where $\rho = 0$.

When a plane wave falls normally on an absorbing medium, all parts of the transmitted wave-front have passed through the same

* *Collected Works*, Vol. III. p. 117.

thickness of the medium. Hence in that case, the wave-front is also a surface of equal amplitude and $\rho = 0$. If the wave-front is normal to the axis of x, so that $\sin\theta = \sin\alpha = 0$, (5) becomes

$$R = R_0 e^{-\frac{2\pi}{\lambda}\kappa_0 x} \cos 2\pi \frac{\nu_0}{\lambda}(x - vt) \quad \text{..............(11).}$$

As ν_0, κ_0 may be calculated from (10), and in this case $v = V/\nu_0$, all quantities are determined. κ_0 is called the coefficient of "extinction."

When the light falls obliquely on the surface the planes of equal amplitude remain parallel to the surface but the planes of equal phase become inclined to the surface at an angle ρ which is the angle of refraction. The expression for the disturbance becomes

$$R = R_0 e^{-\frac{2\pi}{\lambda}\kappa x} \cos 2\pi \frac{\nu}{\lambda}(x\cos\rho + y\sin\rho - vt).$$

Both κ and ν depend on the angle ρ, but (10) shows that $\nu^2 - \kappa^2$ and $\nu\kappa\cos\rho$ are constant, which gives the relations

$$\left.\begin{aligned} \nu^2 - \kappa^2 &= \nu_0^2 - \kappa_0^2 \\ \nu\kappa\cos\rho &= \nu_0\kappa_0 \end{aligned}\right\} \quad \text{.....................(12).}$$

Ketteler, who first realized the importance of these equations, called them the principal equations of wave propagation in absorbing media. The optical distance between two points on the same wave normal, at a distance d apart, is $\frac{2\pi}{\lambda}\nu d$, and we may call ν the coefficient of optical length.

We obtain from (10)

$$\left.\begin{aligned} \nu_0^2 - \kappa_0^2 &= GV^2 \\ 2\nu_0\kappa_0 &= FV^2/\omega \end{aligned}\right\} \quad \text{........................(13).}$$

148. The laws of refraction in absorbing media. Let the disturbance of the incident light be proportional to

$$e^{i(ax+by-\omega t)};$$

the refracted disturbance will be proportional to

$$e^{i(a_1x+by-\omega t)},$$

the identity of the coefficients b and ω on the two sides of the separating surface being proved, as in the case of transparent media. If λ/ν is the wave-length of the transmitted light and ρ the angle of refraction

$$\sin\theta = \nu\sin\rho,$$

or

$$\nu^2\cos^2\rho = \nu^2 - \sin^2\theta.$$

Equations (12) may now be written

$$\nu^2 - \kappa^2 = \nu_0^2 - \kappa_0^2,$$

$$\nu^2\kappa^2 - \kappa^2\sin^2\theta = \nu_0^2\kappa_0^2.$$

Solving these we find

$$\left.\begin{aligned} 2\nu^2 &= \sqrt{(\nu_0^2 - \kappa_0^2 - \sin^2\theta)^2 + 4\nu_0^2\kappa_0^2} + (\nu_0^2 - \kappa_0^2 + \sin^2\theta) \\ 2\kappa^2 &= \sqrt{(\nu_0^2 - \kappa_0^2 - \sin^2\theta)^2 + 4\nu_0^2\kappa_0^2} - (\nu_0^2 - \kappa_0^2 - \sin^2\theta) \end{aligned}\right\} \ldots(14).$$

These equations represent Ketteler's law of refraction for absorbing media. Having found ν for a particular incidence θ by means of the first of equations (14), $\rho = \sin^{-1}(\nu^{-1}\sin\theta)$ gives the angle of refraction.

In the case of transparent media, the wave velocities in the two media are $c/\sqrt{a^2+b^2}$ and $c/\sqrt{a_1^2+b^2}$ respectively, and the refractive index is

$$\mu = \sqrt{\frac{a_1^2+b^2}{a^2+b^2}} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots(15).$$

We shall take this equation to be a definition of the symbol μ also in the case of opaque media. Though μ is now a complex quantity and has no physical meaning, it is useful as an intermediate variable.

Let the first medium be transparent, so that $a^2+b^2=4\pi^2/\lambda^2$. Applying equation (7) to the second medium, and noting that $\theta-\alpha$ is the angle we now denote by ρ, while a must be replaced by a_1, we find:

$$\begin{aligned} \mu^2 = \frac{a_1^2+b^2}{a^2+b^2} &= (\nu^2-\kappa^2) + 2i\kappa\nu\cos\rho \\ &= \nu_0^2 - \kappa_0^2 + 2i\kappa_0\nu_0, \\ \therefore\ \mu &= \nu_0 + i\kappa_0 \ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (16) \end{aligned}$$

The constants ν_0 and κ_0 are therefore derived directly from μ, being respectively its real and imaginary part.

We associate with μ an unreal angle θ_1 defined by

$$\sin\theta = \mu\sin\theta_1.$$

From this we obtain

$$\begin{aligned} \mu^2\cos^2\theta_1 &= \mu^2 - \sin^2\theta \\ &= \nu^2 - \kappa^2 + 2i\kappa\nu\cos\rho - \sin^2\theta \\ &= \nu^2\cos^2\rho - \kappa^2 + 2i\kappa\nu\cos\rho \\ &= (\nu\cos\rho + i\kappa)^2, \end{aligned}$$

$$\therefore\ \sin\theta\cot\theta_1 = m + i\kappa \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(17),$$

where $$m = \nu\cos\rho \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(18).$$

149. Free and Forced Vibrations. A particle attracted to a fixed centre by a force varying as the distance, was proved in Chapter I. to perform simple periodic oscillations. We may extend the investigation now, by admitting the possibility that the motion of the particle is resisted by some frictional force which acts in proportion to its velocity. The path being rectilinear, the equation of motion is

$$\ddot{z} + 2k\dot{z} + n^2 z = 0,$$

having for solution

$$z = Ae^{-kt}\cos\{\sqrt{n^2 - k^2}\,.\,t - \alpha\} \quad \text{.................(19)},$$

where A and α are two constants of integration to be determined by the initial conditions of motion. It will be noticed that the interval between two successive maxima of displacement is increased by the friction and that each maximum is smaller than its predecessor. The motion is no longer simply periodic.

If the friction be so great that $n^2 - k^2$ is negative, the form of the solution alters and the motion becomes "aperiodic," but for our present purpose, we may leave this case out of account. When k is small, the period may be expressed in terms of a series

$$\frac{2\pi}{\sqrt{n^2 - k^2}} = 2\pi\left[\frac{1}{n} + \frac{k^2}{2n^3} + \dots\right],$$

which shows that the most important term involving k depends on its square, so that even though we may take account of effects depending on the first power of k, the period is not affected by friction if we may neglect the second power. Equation (19) represents the motion which the particle assumes when unacted on by external forces, and is therefore called the *free* vibration.

Let the same particle be now subject to an additional periodic force of period $2\pi/\omega$. Its equation of motion becomes

$$\ddot{z} + 2k\dot{z} + n^2 z = E\cos\omega t,$$

where, m being the mass, $E\cos\omega t/m$ is the force. The complete solution now is

$$z = \frac{E\sin\epsilon}{2\omega k}\cos(\omega t - \epsilon) + Ae^{-kt}\cos\{\sqrt{n^2 - k^2}\,.\,t - \alpha\} \quad \text{...(20)},$$

where $$\tan\epsilon = \frac{2\omega k}{n^2 - \omega^2} \quad \text{...(21)}.$$

The second term represents the free vibration which gradually dies out, leaving permanently the "forced" vibration which is represented by

$$z = \frac{E\sin\epsilon}{2\omega k}\cos(\omega t - \epsilon) \quad \text{....................(22)},$$

and which must now be investigated somewhat more closely.

If $n > \omega$, *i.e.* if the forced period be greater than the natural period, ϵ lies in the first quadrant, and the forced vibration is, as regards phase, behind the force. If, on the other hand, $n < \omega$, the forced vibration is accelerated as compared with the force.

If the forced and free vibrations have the same period, $n = \omega$ and

$$z = \frac{E}{2\omega k} \sin \omega t.$$

Here the motion is a quarter of a period behind the force and the amplitude becomes very great for small values of k.

If there is very little friction, we may put, neglecting higher powers,

$$\epsilon = \tan \epsilon = \sin \epsilon = \frac{2\omega k}{n^2 - \omega^2},$$

and the equation of motion is

$$z = \frac{E}{n^2 - \omega^2} \cos \omega \left(t - \frac{2k}{n^2 - \omega^2}\right) \quad \text{.............. (22}a\text{)}.$$

The friction now only affects the phase. For vanishing k, the phase is in complete agreement with that of the force when $n > \omega$ and in complete disagreement when $n < \omega$.

As a suggestive example of motion due to periodic forces, we may work out the case of one pendulum having mass m, and length l, suspended from another pendulum of mass M and length L. For the equations of motion of m, we have, neglecting friction and confining ourselves to small motions,

$$\ddot{x}_1 + \frac{g}{l}(x_1 - x) = 0,$$

where x_1 and x are the displacements of m and M respectively.

If $2\pi/n_1$ be the free period of m, when M is stationary, the equation may be written

$$\ddot{x}_1 + n_1^2 (x_1 - x) = 0 \quad \text{......................(23)}.$$

To form the equations of motion of M, we may take the tension of the upper and lower strings to be $(M+m)g$ and mg respectively. Hence writing α for the ratio of the masses, and $2\pi/n$ for the free period of M when the lower string is not attached, the equation of motion becomes

$$\ddot{x} + n^2 (1 + \alpha) x + n_1^2 \alpha (x - x_1) = 0 \quad \text{..............(24)}.$$

We easily obtain a particular solution of these equations applying to the case that both pendulums perform simple oscillations having the same period. Writing for this purpose

$$\left.\begin{aligned} x &= a \cos \omega t \\ x_1 &= ra \cos (\omega t - \epsilon) \end{aligned}\right\} \quad \text{....................(25)}.$$

We see at once by substitution in (23) and (24) that $\epsilon = 0$.

The same substitution then leads to

$$-\omega^2 + n^2 (1 + \alpha) + n_1^2 \alpha (1 - r) = 0,$$

and

$$-r\omega^2 + n_1^2 (r - 1) = 0.$$

Separating the unknown quantities r and ω, we find:

$$\left.\begin{aligned} r &= \frac{n_1^2}{n_1^2 - \omega^2} \\ \text{and} \qquad n^2 - \omega^2 &= a\,\frac{n^2\omega^2 + n_1^2\omega^2 - n^2 n_1^2}{n_1^2 - \omega^2} \end{aligned}\right\} \qquad \ldots\ldots\ldots\ldots\ldots(26).$$

There are two values of ω^2 which satisfy this equation and the most general case of the motion of the combined pendulum is obtained by superposing oscillations of both kinds. If a be small, the second of equations (26) shows that ω is either nearly equal to n or to n_1.

We then find for ω^2 to the first order of magnitude in a, either

$$\omega^2 = n^2\left(1 + a\frac{n^2}{n^2 - n_1^2}\right)$$

or

$$\omega^2 = n_1^2\left(1 - a\frac{n_1^2}{n^2 - n_1^2}\right).$$

Both equations show that the common period $2\pi/\omega$ is not intermediate between the two periods $2\pi/n$ and $2\pi/n_1$ which the pendulums possess separately by virtue of their length.

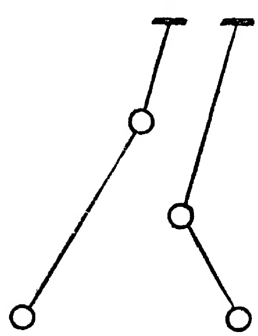


Fig. 170.

If we consider the effect of the lower pendulum simply as a disturbance of that of the upper and heavier one, the first equation may be supposed to hold and it is seen that if the natural period of the upper pendulum is greater than that of the lower, it is made to go still more slowly by the attachment. If a be small, r approaches in the two cases the values $n_1^2/(n_1^2 - n^2)$ and $(n^2 - n_1^2)/an_1^2$. In the first case there is agreement or disagreement of phase according as the upper pendulum has the longer or shorter period. The reverse holds when the combined oscillation has a period which lies near that of the lower one: this case is illustrated in Fig. 170.

150. Passage of light through a responsive medium. We now consider light to pass through a medium, the particles of which are subject to forces, capable of giving rise to free vibrations of definite periods. We consider plane waves propagated in the x direction, the displacements being in the z direction. In order to obtain a simply periodic motion for the free vibrations of the particles, we may imagine each to be attracted to a fixed centre by a force varying as the distance. This centre of force we take to form part of the medium to which it is rigidly attached. If ζ be the displacement of the medium, and ζ_1 that of the particle, the equation of motion of the particle is

$$\ddot{\zeta}_1 + n^2(\zeta_1 - \zeta) = 0 \quad \ldots\ldots\ldots\ldots\ldots\ldots(27).$$

If ρ_1 is the mass of the particle, $n^2\rho_1$ is the force of attraction at unit distance from the centre of force. The reaction of that force has to be taken into account in forming the equations of motion of the medium. At each centre, the medium is acted on by a force $n^2\rho_1(\zeta_1-\zeta)$, and if there are a great many particles within the distance of a wave-length, we may average up the effects and imagine all the forces to be uniformly distributed. Let ρ be the inertia of that portion of the medium which contains on the average one and only one particle. Then the equation for the propagation of the wave is:

$$\ddot{\zeta}+\beta n^2(\zeta-\zeta_1)-V^2\frac{d^2\zeta}{dx^2}=0 \quad\text{.................(28)},$$

where we have written $\beta=\rho_1/\rho$ and V stands for the velocity of propagation when there are no particles or when $n=0$.

If the wave is of the simple periodic type

$$\zeta=\cos(ax-\omega t) \quad\text{..........................(29)}$$

and if the motion has continued without disturbance for a sufficiently long time for the free vibrations of the particle to have died out, their position is expressed by

$$\zeta_1=r\cos(ax-\omega t) \quad\text{.......................(30)}.$$

x is a parameter which is constant for each particle, but varies from particle to particle. By substituting (29) and (30) into (28) and (27), two equations to determine a and r are obtained:

$$-\omega^2 r+n^2(r-1)=0,$$

$$-\omega^2-\beta n^2(r-1)+V^2a^2=0.$$

The first equation gives

$$r=\frac{n^2}{n^2-\omega^2} \quad\text{..........................(31)},$$

and the second
$$\frac{V^2a^2}{\omega^2}=1+\frac{\beta n^2}{n^2-\omega^2}.$$

ω/a is the velocity (v) of transmission of the wave having a frequency $\omega/2\pi$, so that finally

$$\frac{1}{v^2}=\frac{1}{V^2}\left(1+\frac{\beta n^2}{n^2-\omega^2}\right) \quad\text{....................(32)}.$$

The frequency of the free vibration is $n/2\pi$.

This is Sellmeyer's equation, by means of which he first showed that the velocity of light must depend on the periods of free vibration of the molecules embedded in the æther.

151. General investigation of the effect of a responsive medium. It will be useful to introduce here a more general investigation, which we shall base on the electromagnetic theory.

In Art. 137 we had expressed the total current as the sum of a polarization or displacement current and the conduction current. To this, we may now add the convection currents. If there are N positive electrons in unit volume, each carrying a charge e and moving with velocity $\dot{\zeta}_1$ in the z direction, then $Ne\dot{\zeta}_1$ is the z component of the convection current, and to this we must add the convection current of negative electricity $-Ne\dot{\zeta}_2$. We may include both currents in the expression $Ne\dot{\zeta}$ if $\dot{\zeta}$ denote the relative velocity of positive and negative electricity. The conduction current is also due to the convection of electrons, but we leave it in the form CE, because we want to distinguish between the current subject to ohmic resistance which forms a system depending only on one variable, and that which is due to oscillations of electric charges within the molecule. Confining therefore $\dot{\zeta}$ to the velocity of these charges, we have for the z component of the total current in place of (17) Chapter X.,

$$w = \frac{1}{4\pi} K \frac{dR}{dt} + CR + Ne\dot{\zeta} \quad \text{..................(33)}.$$

The last of equations (12) Chapter X. gives with the help of (13) and putting the magnetic permeability equal to one:

$$4\pi \frac{dw}{dt} = \frac{d}{dy}\left(\frac{dR}{dy} - \frac{dQ}{dz}\right) - \frac{d}{dx}\left(\frac{dP}{dz} - \frac{dR}{dx}\right)$$

$$= \nabla^2 R - \frac{d}{dz}\left(\frac{dP}{dx} + \frac{dQ}{dy} + \frac{dR}{dz}\right) \quad \text{..............(34)}.$$

The last term vanishes in isotropic media, and we may in that case, eliminating w, write for the equation of motion in the z direction

$$\frac{d}{dt}\left(K\frac{dR}{dt} + 4\pi CR + 4\pi Ne\dot{\zeta}\right) = \nabla^2 R \quad \text{...........(35)}.$$

Always assuming the disturbance to be simply periodic, the displacement ζ may be divided into two portions, one of which is in phase with R and the other in phase with dR/dt. Writing therefore

$$4\pi\zeta = AR - \frac{B}{\omega}\frac{dR}{dt} \quad \text{..........................(36)},$$

and

$$4\pi\dot{\zeta} = A\frac{dR}{dt} - \frac{B}{\omega}\frac{d^2R}{dt^2}$$

$$= A\frac{dR}{dt} + B\omega R \quad \text{.......................(37)},$$

(35) becomes

$$(K + NAe)\frac{d^2R}{dt^2} + (4\pi C + \omega NBe)\frac{dR}{dt} = \nabla^2 R.$$

Comparing this with (1), we see that the results of Art. 147 may be applied to this case, putting

$$G = K + NAe,$$

$$F = 4\pi C + \omega NBe.$$

Hence by (13) $$\nu^2 - \kappa^2 = (K + NAe)\, V^2 \quad \text{......................(38)},$$
$$2\nu\kappa \cos\rho = (2C\lambda + NBVe)\, V \quad \text{..................(39)}.$$

If the specific inductive capacity of the intermolecular space is the same as that of empty space, $KV^2 = 1$, and (38) is replaced by
$$\nu^2 - \kappa^2 = 1 + NAeV^2 \quad \text{......................(40)}.$$

152. Wave velocity in a responsive medium according to the electromagnetic theory. P. Drude was the first to apply the electromagnetic theory to the explanation of dispersion based on the principles of sympathetic vibrations. It is now generally accepted that each molecule contains a number of electrons, and that each electron consists of a definite electric charge concentrated within a space which must be small compared with the size of the molecule. Every quantity of electricity is made up of these electrons just as any ordinary substance is made up of atoms. It is therefore correct to speak of an electron as an "atom" of electricity. A moving electron represents kinetic energy which exists in the medium surrounding the electron in virtue of the magnetic field established by the motion. This energy is proportional to the square of the velocity, and may be expressed therefore in the form $\frac{1}{2}\rho v^2$ if v be the velocity of the electron. The quantity ρ we may call the apparent inertia of the electron. Students must however guard against being misled into the error of believing that if there are several electrons near each other, their total kinetic energy may be written down as the sum of their separate kinetic energies. That total energy contains products of the form $v_1 v_2$, where v_2 and v_1 refer to distinct electrons. While dealing with a single electron we are, however, justified in applying to it the ordinary laws of dynamics substituting ρ for its mass. If we therefore consider that the incident vibrations of light excite the sympathetic motion of a single electron in a molecule, we may write for the equation of motion
$$\ddot{\zeta} + n^2\zeta - \frac{eR}{\rho} = 0.$$

In forming this equation, we have imagined the electron to be acted on by a centre of force varying as the distance, the force being $n^2\rho$ at unit distance, while e is the charge of the electron and R the external electric force acting on it. If R varies as $e^{i\omega t}$ and the free vibration has died out, so that the period of ζ is that of the incident force, we may deduce
$$\zeta = \frac{eR}{\rho\,(n^2 - \omega^2)}.$$

By comparison with (36) we see that in the present case $B = 0$, and
$$A = \frac{4\pi e}{\rho\,(n^2 - \omega^2)} \quad \text{..........................(41)}.$$

If the medium is a non-conductor, (39) shows that $\kappa = 0$ and ν in that case is equal to the refractive index μ. We therefore finally obtain from (40)

$$\mu^2 = 1 + \frac{4\pi e^2 N V^2}{\rho(n^2 - \omega^2)} \quad \text{.......................(42).}$$

If the charge e is measured in the electrostatic system, we may leave out the factor V^2 in the second term of the right-hand side, and the equation is then identical with that given by Drude*. The dimensions of the equation are easily checked, as e^2/ρ is of the dimension of a length.

The manner in which the refractive index changes with a change in the wave-length of the incident light, is easily recognized by inspection of (42). For long waves, ω is small, and in the limit when $\omega = 0$:

$$\mu^2 = 1 + \frac{4\pi e^2 N V^2}{\rho n^2}.$$

With increasing ω the refractive index increases until $n = \omega$. At that point there is a discontinuity, μ^2 suddenly changing from $+\infty$ to $-\infty$.

For a definite value of ω larger than n, μ is zero and afterwards μ increases once more and approaches unit value for infinitely short periods. If we introduce K' the specific inductive capacity of the medium which must coincide with the value of μ^2 for $\omega = 0$, we may write

$$\mu^2 = K' + \frac{4\pi e^2 N V^2 \omega^2}{\rho (n^2 - \omega^2) n^2} \quad \text{.....................(43).}$$

Fig. 171 represents the curve $y = \dfrac{1}{1 - x^2}$. By a change of scale, and a displacement of the horizontal axis, the curve may be made to coincide with that which gives the relation between μ^2 and ω. In order to establish agreement with the subsequent curves in which small values of ω are placed on the right, the positive axis of x in the figure is drawn towards the left.

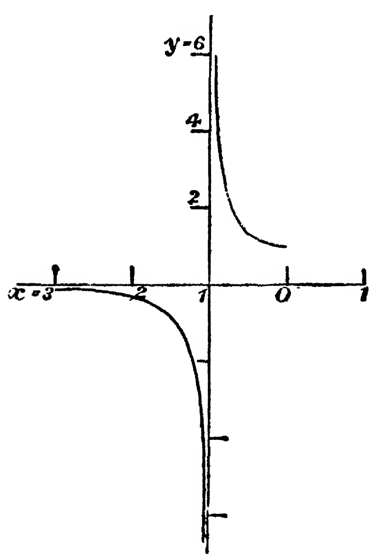


Fig. 171.

The above investigation gives the essential features of the theory of dispersion in its simplest form, and may be extended so as to approach more nearly the actual conditions. Within each molecule there are a number of free periods of vibration as is shown by spectroscopic

* *Lehrbuch der Optik*, Chapter v.

observation. It is necessary to consider therefore a connected system of electrons, the state of motion for each free period being defined by one variable. Assuming these periods to be independent of each other, let ψ_1, ψ_2, etc. be the variables. For small changes $\delta\psi_1$, $\delta\psi_2$, of these variables, the total work done may be expressed in the form

$$\Psi_1\delta\psi_1 + \Psi_1\delta\psi_2 + \ldots = 0.$$

If the displacements from the stable configuration are small, an equation of motion holds for each variable of the form

$$\ddot{\psi}_m + n_m^2\psi_m = \Psi_m{}^*,$$

where Ψ_m is the generalized component of force. The frequency of the free period being $n_m/2\pi$, and the motion being assumed to be proportional to $e^{i\omega t}$, the equation may be written

$$(n_m^2 - \omega^2)\,\psi_m = \Psi_m.$$

If the external electric force is $Re^{i\omega t}$, we have for the different variables

$$(n_1^2 - \omega^2)\,\psi_1 = \Psi_1 = A_1R,$$
$$(n_2^2 - \omega^2)\,\psi_2 = \Psi_2 = A_2R,$$

etc. where A_1 and A_2 are constants. Linear relationships must hold between the displacements ζ_1, ζ_2, etc. of the electrons and ψ_1, ψ_2, so that for the total current within each molecule, we have

$$e(\dot{\zeta}_1 + \dot{\zeta}_2 + \dot{\zeta}_3 + \ldots) = e(a_1\dot{\psi}_1 + a_2\dot{\psi}_2 + \ldots)$$
$$= e\frac{dR}{dt}\left(\frac{a_1A_1}{n_1^2 - \omega^2} + \frac{a_2A_2}{n_2^2 - \omega^2} + \ldots\right) \ldots\ldots\ldots(44).$$

Hence A in (37) becomes

$$A = \Sigma\frac{a_mA_m}{n_m^2 - \omega^2} \;\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(45).$$

a_1, a_2, are a set of constants, the connexion between which and A_1, A_2 etc. it is not necessary to discuss here †.

From the equation defining the electric current, we may, as in the simple case which has been treated in detail, derive the expression corresponding to (42) which now takes the form

$$\mu^2 = 1 + \frac{\beta_1}{n_1^2 - \omega^2} + \frac{\beta_2}{n_2^2 - \omega^2} + \ldots \;\ldots\ldots\ldots\ldots(46),$$

where β_1, β_2, etc. are numerical constants defining the dispersion of the medium. For waves of infinitely long periods, we have

$$\mu_\infty{}^2 = K' = 1 + \frac{\beta_1}{n_1^2} + \frac{\beta_2}{n_2^2} + \frac{\beta_3}{n_3^2} + \ldots \;\ldots\ldots\ldots(47).$$

If there is only one period of vibration, β_1 may be determined from the observed inductive capacity, and if K' is known, the refractive index for all waves is completely determined.

* Lord Rayleigh, *Theory of Sound*, Vol. I Chapter IV.

† A general theory of optical dispersion was first given by Larmor: *Phil. Trans.* CXC. p. 236 (1898).

By using (47) equation (46) may be written, changing the constants,

$$\mu^2 = K' + \Sigma \frac{M_m}{\lambda^2 - \lambda_m^2} \quad \text{.....................(48)},$$

where λ is the wave-length in vacuo to which ω applies, and λ_m is the wave-length, also measured in vacuo, which corresponds to the free vibration of the molecules.

153. Dispersion in transparent media. If the range of spectrum considered is far removed from any of the free periods of the molecule, the dispersion formula may conveniently be expressed differently. If the region of resonance lies in the ultra-violet, we may expand in terms of a series proceeding by λ_m/λ and thus find

$$\mu = K' + \frac{A_1}{\lambda^2} + \frac{A_2}{\lambda^4} + \dots \quad \text{..................(49)},$$

where $A_1 = \Sigma M_m$; $A_2 = \Sigma M_m \lambda_m^2$; $A_p = \Sigma M_m \lambda_m^{2p-2}$.

Equation (49) is known by the name of Cauchy's formula, but was deduced by Cauchy in quite a different manner.

If the region of resonance is in the infra-red, we may express the series in terms which proceed by $\frac{\lambda}{\lambda_m}$ and thus obtain

$$\mu^2 = K' - B_0 - (B_1\lambda^2 + B_2\lambda^4 + \dots) \quad \text{..............(50)},$$

where $B_0 = \Sigma \frac{M_m}{\lambda_m^2}$; $B_1 = \Sigma \frac{M_m}{\lambda_m^4}$; $B_2 = \Sigma \frac{M_m}{\lambda_m^6}$; $B_n = \Sigma \frac{M_m}{\lambda_m^{2p+2}}$.

In the case of many substances Cauchy's formula does not give a sufficient representation of the actual dispersion without the addition of a negative term proportional to λ^2. This fact which has been clearly established by Ketteler suggests that though the dispersion in the visible part is mainly regulated by ultra-violet resonance, it is also to some extent influenced by free periods lying in the infra-red. Assuming for the sake of simplicity one infra-red and one ultra-violet free period, having wave-lengths λ_r and λ_v respectively, equation (48) becomes

$$\mu^2 = K' + \frac{M_v}{\lambda^2 - \lambda_v^2} + \frac{M_r}{\lambda_r^2 - \lambda^2} \quad \text{..................(51)}.$$

This equation has been tested over a long range of wave-lengths for rock-salt, sylvin and fluorspar, and the agreement arrived at is sufficient to show that in its essential points, the present theory is correct, and that refraction is a consequence of the forced vibration of the molecules, which respond strongly to the periodic impulses of those waves which are in sympathy with its periods of free vibration. These experiments, which are fundamental to the theory of refraction, have been made possible by the beautiful device of H. Rubens and E. F. Nichols*, which enabled them to obtain fairly homogeneous radiations of large wave-lengths by multiple reflexion. The success of the method itself is an excellent confirmation of the above theory.

According to Sellmeyer's equation, the refractive index is infinite for periods equal to those of the free vibrations; light of such periods is totally reflected. By successive reflexions from a number of surfaces, all wave-lengths are eliminated except those for which there is approximately total reflexion. It was found in this manner that with quartz, wave-lengths of 20·75 and 8·25 mikrons, and with fluorspar a wave-length of 23·7, could be obtained.

The refractive index of quartz is represented with considerable accuracy by the formula

$$\mu_2 = K' + \frac{M_v}{\lambda^2 - \lambda_v^2} - \frac{M_r}{\lambda_r^2 - \lambda^2} - \frac{M_s}{\lambda_s^2 - \lambda^2},$$

in which λ_r and λ_s are directly determined by observation.

Rubens and Nichols also determined the refractive indices of rock-salt and sylvin for the wave-lengths 20·75μ and 8·85μ and hence could deduce an equation to represent the dispersion of these two substances through a wide range. In the following Table, I have collected the constants of the substances used by the authors, adding Paschen's† numbers for fluorspar.

TABLE X.

	Quartz	Fluorspar	Rock-salt	Sylvin
M_v	·01065	·00612	·01850	·0150
M_r	44·224	5099	8977	10747
M_s	713·55	—	—	—
λ_v	·1031	·09425	·127	·153
λ_r	8·85	35·47	56·12	67·21
λ_s	20·75			
K'	4·5788	6·0910	5·179	4·553
K	4·55	6·8	5·85	4·94

All wave-lengths are given in mikrons, *i.e.* in 10^{-4} cms. As has been stated, the resonance periods of quartz have been derived from observation, the others are calculated from the dispersion

* *Wied. Ann.* LX. p. 418 (1897).

† *Wied. Ann.* LIII. p. 812.

formula. A good confirmation was subsequently obtained by H. Rubens and E. Aschkinass* in the direct determination of the resonance region in rock-salt and sylvin, though the observed free periods were not found to coincide as much as might have been wished with those derived from the dispersion formula. The wave-lengths for total reflexion were measured to be 51·2 and 61·1 instead of 56·1 and 67·2 as given in the table. There is still less agreement in the case of fluorspar, the wave-length best fitting the observation being 35·47, while the region of total reflexion lies at 23·7. The discrepancy may be due to the fact that as in quartz, fluorspar has a second region of total reflexion in the infra-red.

The next remark called for by the inspection of the table is connected with the relative small values of the constants M in the ultra-violet term. This must be due to the comparative smallness of the resonance for short periods. There is a gradual increase of the value of M for diminishing values of the resonance period. This increase is not very uniform, but is such that in general it is more rapid than the increase in the square of the wave-length at which resonance takes place. A closer investigation of this point seems called for, but it would be necessary for the purpose to take account of the molecular volumes of the different substances. If the refractive index for infinitely short waves is one, as required by (46), equation (48) shows that the constants should satisfy the condition

$$\Sigma \frac{M_m}{\lambda_m^2} = K' - 1.$$

This relation is only approximately fulfilled by the numbers given in the Table, but its complete verification was not to be expected considering that there are probably unknown regions of resonance in the ultra-violet.

The constant K' should, according to theory, be equal to the specific inductive capacity of the medium; the last two rows of the table show that though the present agreement is not by any means perfect, its power to represent the facts is a considerable stage in advance of the older theories which gave Cauchy's formula. (Art. 139.)

154. Extension of the theory. Our theory requires extension in two directions. Sellmeyer's equation

$$\mu^2 - 1 = \Sigma \frac{\beta_m n_m^2}{n_m^2 - \omega^2}$$

gives infinitely large values of μ whenever the period of the incident

* *Wied. Ann.* LXV. p. 241 (1898).

light coincides with one of the free periods of vibration. This is a consequence of the infinite amplitude of the forced vibration, as it appears *e.g.* in equation (22*a*).

These infinite amplitudes may be avoided by the introduction of a frictional term retarding the free vibrations as in the case treated in Art. 147. Real friction is not admissible in the treatment of molecular vibrations, but as there is loss of energy due to radiation, their must be some retarding force, which is in phase with the velocity and which recent researches have shown to depend on the third differential coefficient of the displacement. Its effect will be the same as that of a frictional force. Confining ourselves for simplicity to a single variable, we may write for the displacement of the electrons in the molecule according to (22)

$$\zeta = \frac{R_0 e \sin\epsilon}{2\rho\omega k}\cos(\omega t - \epsilon),$$

where $R = R_0 \cos\omega t$ represents the electric force due to the incident light and ρ has the same meaning as in Art. 152.

Introducing R in place of R_0 the equation becomes

$$\zeta = \frac{e\sin 2\epsilon}{4\rho\omega k}R - \frac{e\sin^2\epsilon}{2\rho\omega^2 k}\frac{dR}{dt},$$

where by (21) $\quad \tan\epsilon = \dfrac{2\omega k}{n^2-\omega^2}.$

Hence by comparison with (36)

$$\begin{aligned} A &= \frac{\pi e\sin 2\epsilon}{\rho\omega k} = \frac{2\pi e}{\rho\omega k}\frac{\tan\epsilon}{1+\tan^2\epsilon} \\ &= \frac{4\pi e}{\rho}\frac{n^2-\omega^2}{4\omega^2k^2+(n^2-\omega^2)^2} \quad\text{......................(52),} \end{aligned}$$

and

$$\begin{aligned} B &= \frac{2\pi e\sin^2\epsilon}{\rho\omega k} = \frac{2\pi e}{\rho}\frac{\tan^2\epsilon}{1+\tan^2\epsilon} \\ &= \frac{2\pi e}{\rho}\frac{4\omega k}{4\omega^2k^2+(n^2-\omega^2)^2}. \end{aligned}$$

Hence (39) and (40) become, taking account of (12), and putting $C=0$,

$$\left.\begin{aligned} \nu_0^2 - \kappa_0^2 &= 1 + 4\pi Ne^2V^2\frac{n^2-\omega^2}{\rho\{4\omega^2k^2+(n^2-\omega^2)^2\}} \\ \nu_0\kappa_0 &= \phantom{1+{}} 2\pi Ne^2V^2\frac{2\omega k}{\rho\{4\omega^2k^2+(n^2-\omega^2)^2\}} \end{aligned}\right\} \quad\text{.........(53).}$$

If κ_0 be small, so that κ_0^2 may be neglected, ν_0 becomes equal to the refractive index of the substance, which then refracts according to the sine law. The introduction of k has got rid of the infinite value of ν_0, but the value of k will be shown in Chapter XIII. to be too small to be the cause of the observed absorption phenomena.

155. Finite range of Free Vibrations. Solids and liquids exhibit absorption bands extending frequently with varying intensity over a considerable range of the spectrum. In order to include this case in our theory we may imagine that the free vibrations are not confined to definite periods. The summation in (46) must then be replaced by an integration, and we write therefore:

$$\mu^2 - 1 = \int_{n_1}^{n_2} \frac{\beta dn}{n^2 - \omega^2}.$$

Here β may be a function of n. Assuming it to be constant, we find on integration

$$\mu^2 - 1 = \frac{\beta}{2\omega} \log \left| \frac{(n_2 - \omega)(n_1 + \omega)}{(n_2 + \omega)(n_1 - \omega)} \right| \ldots\ldots\ldots\ldots\ldots (54),$$

where the absolute value of the fraction, the logarithm of which occurs in the expression, is to be taken. The square of the refractive index is infinitely large on the positive side for $\omega = n_1$, and infinitely large on the negative side for $\omega = n_2$, n_2 belonging to the higher frequency. The region of ω for which μ^2 is negative includes that range of waves which cannot enter the medium. It is bounded on the red side by the value of ω for which

$$\log \frac{(\omega - n_1)(n_2 + \omega)}{(n_2 - \omega)(n_1 + \omega)} = \frac{2\omega}{\beta},$$

and on the side of higher frequency by the value of ω for which

$$\log \frac{(n_1 - \omega)(n_2 + \omega)}{(n_1 + \omega)(n_2 - \omega)} = \frac{2\omega}{\beta}.$$

The infinity of the refractive index at the lower frequency edge of the absorption is avoided if the intensity of the absorption band is assumed to diminish gradually to zero on both sides instead of beginning and ending abruptly. As a simple example we may take the case that the absorption between n_1 and n_2 is equal to $\beta(n - n_1)(n_2 - n)$. The expression for $\mu^2 - 1$ then becomes

$$\begin{aligned}
\mu^2 - 1 &= \int_{n_1}^{n_2} \frac{\beta(n - n_1)(n_2 - n)}{n^2 - \omega^2} dn \\
&= \beta \int_{n_1}^{n_2} \left\{ \frac{(n_1 + \omega)(n_2 + \omega)}{2\omega(n + \omega)} - \frac{(n_1 - \omega)(n_2 - \omega)}{2\omega(n - \omega)} - 1 \right\} dn \\
&= \beta \left\{ \frac{(n_1 + \omega)(n_2 + \omega)}{2\omega} \log \frac{n_2 + \omega}{n_1 + \omega} \right. \\
&\qquad \left. - \frac{(n_1 - \omega)(n_2 - \omega)}{2\omega} \log \left| \frac{n_2 - \omega}{n_1 - \omega} \right| + n_1 - n_2 \right\} \ldots\ldots\ldots (55).
\end{aligned}$$

In the second term the sign has to be chosen so as to give always a positive value to the fraction. At the edges of the band, we have for $\omega = n_1$:

$$\mu^2 - 1 = \beta \left\{(n_1 + n_2) \log \frac{n_1 + n_2}{2n_1} - (n_2 - n_1)\right\};$$

for $\omega = n_2$:

$$\mu^2 - 1 = \beta \left\{(n_1 + n_2) \log \frac{2n_2}{n_1 + n_2} - (n_2 - n_1)\right\}.$$

156. Absorption. The gradual extinction of a wave of light as it traverses a medium does not necessarily imply absorption in the proper meaning of the term, which should only be applied when the medium retains the energy which it has abstracted converting it into heat or other forms of energy, such as chemical changes, that is to say, when there is a degradation of energy. We have discussed the effects of the free periods of the electrons which form part of the molecular structure of the medium; the electric forces of the wave as it passes produce steady oscillations of these electrons which in their turn affect the wave, diminishing its amplitude and altering its velocity. But the steady state is soon reached, and once established the electrons gain no longer in energy, and what they take away from the wave they give up to the medium as independently radiating centres. This, properly speaking, is scattering. It would be wrong to describe a change of refractive index as an effect of absorption: it is an effect of scattering.

True absorption is probably due partly to molecular encounters and partly to molecular motion set up by pressure of light. One effect of molecular encounters will be to restore the free vibration included in (20) but left out of account in the subsequent treatment of the problem. If the incident light extends over a finite range of the spectrum we must multiply the first term of that equation by $d\omega$ and integrate. Introducing the value of ϵ in the first term we obtain:

$$z = \int_{O_1}^{\omega_2} E d\omega \left[(n^2 - \omega^2)^2 + 4\omega^2 k^2\right] \cos(\omega t - \epsilon) \\ + \Sigma A_s e^{-n(t - t_s)} \sin \sqrt{n^2 - k^2}\,(t - t_s).$$

The expression under summation indicates that at time t_s the velocity of the particle has altered by a quantity proportional to A_s, the displacement remaining continuous. The two terms, one which through an integration represents the forced vibration due to light, and the second by means of a summation the free vibrations provoked by molecular shocks, have a form which in appearance is different. But it can be shown that optically they are identical*, as may be proved by expressing the free vibrations in terms of Fourier's integral. Both terms represent vibrations which are the more nearly homogeneous the

* Schuster, *Proc. Roy. Soc.* Vol. VIII. p. 248 (1920).

smaller the value of k, and it has already been pointed out that as far as its effects on the refraction of light is concerned k is negligible.

The gradual extinction of light passing through a medium in which scattering preponderates follows a law different from the exponential one which holds in the case of absorption; the reason being that the scattered light is itself thrown backwards and forwards by further scattering. To take a simple case, we may imagine a thick layer of the medium in which a general radiation equal in all directions falls on one surface which we take to be in the plane of YZ. If we take A to be the radiation which falls on a plane parallel to this surface at a distance x from it, proceeding from the negative to the positive side, and $2s$ to be the coefficient of scattering, then a layer of thickness dx will scatter light of total amount $2sAdx$, of which half is thrown forwards and half backwards. Similarly, light of intensity B, which proceeds in the negative direction, will lose by scattering an amount $2sBdx$, which is also divided into two equal halves. We therefore have:

$$\frac{dA}{dx} = \frac{dB}{dx} = s(B - A),$$

from which it follows that:

$$\frac{d(A+B)}{dx} = 2s(B-A); \quad \frac{d(A-B)}{dx} = 0.$$

The second equation shows that $A - B$ is a constant for which we may write R. At the surface from which the light ultimately emerges $B = 0$, and R denotes therefore the intensity of radiation which passes through the layer. The first equation now gives:

$$\frac{dA}{dx} = -sR,$$

and hence $$A = E - sRx,$$

where E is the radiation which enters. If the total thickness is t, A must be equal to R for $x = t$. Hence $R = E/(1 + st)$: and finally:

$$A = E\,[1 + s(t-x)/(1+st)]\,; \; B = Es(t-x)/1 + st.$$

The equation for R shows that the total loss through scattering increases at a much smaller rate with increasing thickness than in the case of absorption. If the layer is sufficiently thick, so that st is large compared with unity, the incident light is entirely thrown back, and the scattering body acts as a total reflector. In that case:

$$A = B = E(1 - xt).$$

The radiations in the two directions are therefore equal when x is small compared with t, and the radiation is then everywhere equal to that which entered the scattering body.

157. Selective refraction. The phenomena which have called forth the theoretical discussions of this Chapter have been grouped together under the name "Anomalous Dispersion." But we are now prepared to say that there is nothing anomalous in the effect of absorption on refraction, and that the ordinary or "normal" dispersion is only a particular case of the "anomalous" one. Under these circumstances the name is misleading, and I therefore introduce the more appropriate one of "Selective Refraction" and "Selective Dispersion."

The experimental illustration of selective refraction is rendered somewhat difficult by the fact that the substances which show the effects are all highly absorbent. With a hollow prism filled with a strong solution of fuchsin or cyanin, it may easily be demonstrated that the red of the spectrum is more refracted than the violet, but dispersion in the immediate neighbourhood of the absorption band is too great to make exact measurements in that region possible. Kundt originated a method of observation which is often employed. The vertical slit of a spectroscope is illuminated by projecting upon it the image of a *horizontal* slit, through which white light is passed. If the horizontal slit be narrow, an almost linear spectrum is seen, running along a horizontal line. The position of this horizontal line may be marked. If now a prism filled with a substance showing selective refraction be interposed between the horizontal slit and its image, the refracting edge of this prism being horizontal and downwards, the line of the observed spectrum will no longer be straight. Were the prism filled with water, the spectrum would run upwards in a curved line from red to violet. A curve running downwards from red to violet would indicate a refractive index diminishing with increasing frequencies. Refractive indices smaller than one, showing a velocity of light greater than that of empty space, would be indicated by displacement below that of the original linear spectrum. For purposes of illustration, and for measurements when the angle of the prism is small, this method is very successful.

The simplest case of selective refraction is shown by sodium vapour, as the absorption is here confined to two regions, each of which covers only a very narrow range of wave-lengths. In other words, the refraction is affected by absorption "lines" as distinguished from absorption bands. The selective refraction of a luminous conical sodium flame was first shown by Kundt, who however did not investigate the specially interesting region which lies between the absorption lines. This has been done by Becquerel. Plate II. Fig. 6 is a reproduction of one of Mr Becquerel's photographs, the red end being to the right. The sodium vapour was used in the form of a luminous flame formed by a special device into a prismatic shape.

The horizontal black line marks the horizontal linear spectrum in its original position. The horizontal portion of the white band, the centre of which is slightly raised above the black line, shows that at a short distance from the double sodium line there is a slight displacement upwards indicating a refractive index somewhat greater than one. The nearly vertical branches of the curve indicate a considerable dispersion close to and between the absorption lines. The course of this dispersion is better studied in Fig. 172, which has been drawn from the measurements given by Mr Henri Becquerel*, upward displacements being approximately proportional to $\mu-1$. It will be seen that in accordance with the previous theory, the refractive index rapidly increases as we approach each absorption line from the red end, and that the light which vibrates just a little more quickly than that corresponding to the absorption band has its velocity increased. Mr Becquerel calculates approximately that the light in close proximity to D_2 and on its violet side has a refractive index of ·9988, so that its velocity is about ·1 % greater than that in empty space. Concentrated solutions of colouring matters exhibit the phenomena of selective refraction, but here the theory is complicated by the fact that the absorption extends over a wide range of wave-lengths. Some of these colouring matters may be shaped into solid prisms of small angle, by means of which the refractive indices for different periods and the coefficients of absorption may be measured. Pflüger† takes a few drops of a concentrated solution of the colouring matters in alcohol, and runs the solution into the two wedge-shaped spaces between a glass plate and a wide glass tube. As the solvent evaporates, it leaves behind a double prism. Amongst many prisms made in this fashion, a few may be found with surfaces sufficiently good to render optical investigation of refractive indices possible. The prisms used by Pflüger had a refracting angle of from 40—130 seconds of arc, and the refractive indices could be determined throughout the absorption band. It is a special merit of Pflüger's

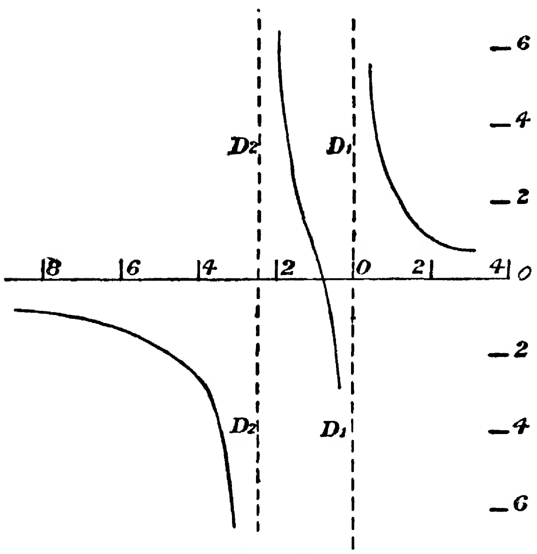


Fig. 172.

* *Comptes Rendus* CXXVIII. p. 145 (1899).

† *Wied. Ann.* LVI. p. 412 (1895) and LXV. p. 173 (1898).

investigations that he determined also the coefficients of absorption for the different wave-lengths. As a thickness of very few wave-lengths is sufficient to extinguish the light, the plates used for the purpose had a thickness of less than half a wave-length. In Figure 173

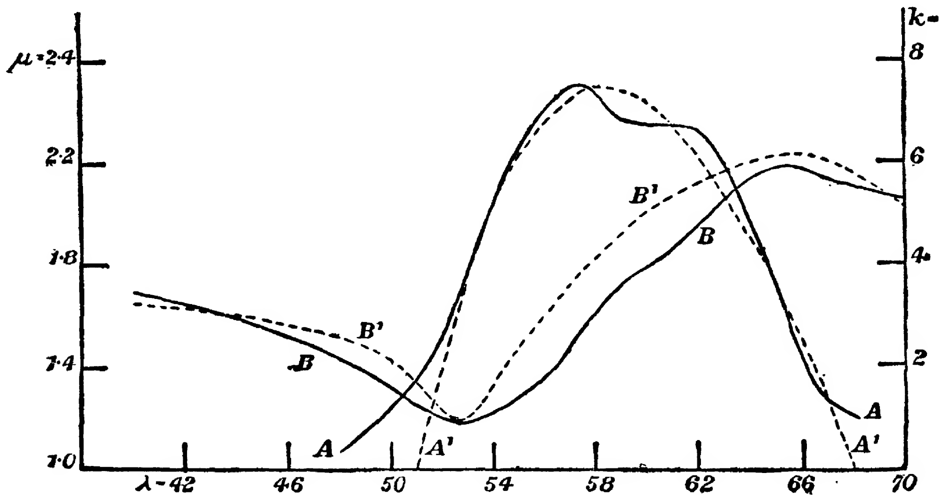


Fig. 173.

Pflüger's curves for cyanin are reproduced, the curves of refractive index and coefficient of extinction being marked B and A respectively. The dotted line A' indicates an assumed absorption curve following the law suggested in Art. 150, the constants being roughly adjusted so as to fit the edges of the absorption band. A second dotted line shows the curve of refraction (B') calculated from equation (55) after substitution of $\nu^2 - \kappa^2$ for μ^2. The value of β was determined so as to give roughly the proper quantity for the difference in the refractive indices near the two edges of the absorption band, and a constant term has been added to represent the effect of infra-red and ultra-violet absorption. The correspondence between calculated and experimental values might be made closer if instead of a constant term, one varying with the wave-length had been taken, but in view of the fact that the assumed law of absorption only approximately represents the facts, it seems unnecessary to seek for a closer agreement of the refraction curves. The more sudden fall and rise of the calculated dispersion curve near the green boundary of the absorption band is due to the fact that the actual absorption curve does not show the rapid increase of absorption indicated by the assumed curve.

158. Metallic Reflexion. We include under the term "metallic" reflexion, all cases in which the greater portion of the incident light is returned, in consequence, as it will appear, of the absorptive power of the medium If the amplitudes of the incident, reflected and refracted light be denoted by 1, r, s, respectively, we may as in Art. 140 write

for the incident wave
$$e^{i(ax+by-\omega t)},$$
for the reflected wave
$$re^{i(-ax+by-\omega t)} \quad\text{.........................}(59),$$
and for the disturbance entering the medium
$$se^{i(a_1x+by-\omega t)} \quad\text{.........................}(60).$$

The surface conditions are the same as for transparent media, hence the previous investigations apply as far as the analytical expressions of r and s are concerned. When the incident light is polarized in the plane of incidence, the amplitude r_s was found to be

$$r_s = \frac{\sin(\theta_1 - \theta)}{\sin(\theta_1 + \theta)} \quad\text{.......................}(61),$$

and for light polarized perpendicularly to the plane of incidence

$$r_p = \frac{\tan(\theta_1 - \theta)}{\tan(\theta_1 + \theta)} \quad\text{.......................}(62).$$

But θ_1 being now complex, r_s and r_p are complex quantities from which we may separately deduce the real amplitude and the change of phase. Writing $r_s = h_s e^{i\delta}$ we see from (59) that h_s is the real amplitude, and δ denotes the change of phase.

The problem now resolves itself into one of algebraic transformation: equations (61) and (62) must be put into the standard form $he^{i\delta}$. If we therefore confine ourselves to calculating the value of h for perpendicular incidence, and μ_0 denote the refractive index of the first medium we have as with transparent bodies:

$$r_s = r_p = \frac{\mu - \mu_0}{\mu + \mu_0}.$$

But from (16)
$$\mu = \nu_0 + i\kappa_0,$$
$$\therefore \quad r_s = he^{i\delta} = \frac{\nu_0 - \mu_0 + i\kappa_0}{\nu_0 + \mu_0 + i\kappa_0} \quad\text{.................}(63).$$

If this quantity be called P, and Q be that obtained from P by reversing the sign of i, the proposition proved in Art. 8 shows that

$$h^2 = PQ\,; \quad \tan\delta = \frac{P - Q}{i(P + Q)}.$$

Hence
$$h^2 = \frac{(\nu_0 - \mu_0)^2 + \kappa_0^2}{(\nu_0 + \mu_0)^2 + \kappa_0^2}\,; \quad \tan\delta = \frac{2\kappa_0\mu_0}{\nu_0^2 + \kappa_0^2 - \mu_0^2} \quad\text{........}(64).$$

It will be noticed that great absorbing power means a large intensity of reflected light, for if κ_0 is large compared with $\nu_0 + 1$, h is nearly 1, and the light is almost totally reflected. The absorbing power therefore is effective not only in transforming the energy that enters, but also in preventing the light from entering. There is also nearly total reflexion when ν_0 is either large or small compared with 1, but this effect is not confined to opaque substances.

When the incidence is oblique, plane polarized light becomes elliptically polarized by reflexion. To determine the constants of the ellipticity, we take the incident beam to be polarized at an angle ϕ to the plane of incidence, and consider separately light of amplitude $\cos\phi$ polarized in that plane and light of amplitude $\sin\phi$ polarized at right angles to it.

We write R_s and R_p for the reflected complex amplitudes of the components. These may be put equal to $r_s \cos\phi$ and $r_p \sin\phi$, if the amplitude of the incident light is unity. To separate the real and complex parts we write $h_s e^{i\delta_1}$ and $h_p e^{i\delta_2}$ for r_s and r_p respectively. We have therefore

$$R_s = h_s \cos\phi \, e^{i\delta_1}, \quad R_p = h_p \sin\phi \, e^{i\delta_2},$$

and writing

$$\tan\phi = h_p/h_s,$$

$$\frac{R_p}{R_s} = \tan\phi \tan\psi e^{i(\delta_2 - \delta_1)}.$$

As δ_1 and δ_2 measure, for the two components, the change of phase at reflexion, the difference in the phase of vibration of the two components after reflexion is equal to $\delta = \delta_2 - \delta_1$. The ratio of the real amplitude is found from the above to be $\tan\phi \tan\psi$. If therefore we restore the plane polarization of the reflected light by accelerating one of the components or retarding the other, the plane of polarization will be inclined to the plane of incidence at an angle χ given by

$$\tan\chi = \tan\phi \tan\psi.$$

The quantities χ and ϕ may be measured, and hence ψ may be found. We must now endeavour to express the optical constants ν_0 and κ_0 in terms of δ and ψ. For this purpose, we have, from (61) and (62),

$$\frac{\cos(\theta_1 + \theta)}{\cos(\theta_1 - \theta)} = \frac{r_p}{r_s} = \frac{h_p e^{i\delta_2}}{h_s e^{i\delta_1}}$$

$$= \tan\psi e^{i\delta}.$$

Also

$$\frac{\cos(\theta_1 + \theta)}{\cos(\theta_1 - \theta)} = \frac{\cos\theta_1 \cos\theta - \sin\theta_1 \sin\theta}{\cos\theta_1 \cos\theta + \sin\theta_1 \sin\theta} = \frac{\cot\theta_1 - \tan\theta}{\cot\theta_1 + \tan\theta}.$$

Hence introducing (17)

$$\tan\psi e^{i\delta} = \frac{m - \tan\theta \sin\theta + i\kappa}{m + \tan\theta \sin\theta + i\kappa} \quad \text{..................(65).}$$

This expression is of the same form as (63), and by subjecting it to the same transformation, we find

$$\tan^2\psi = \frac{(m - \tan\theta \sin\theta)^2 + \kappa^2}{(m + \tan\theta \sin\theta)^2 + \kappa^2}.$$

Applying $$\cos 2\psi = \frac{1-\tan^2\psi}{1+\tan^2\psi},$$
this reduces to
$$\cos 2\psi = \frac{2m\cos\theta\sin^2\theta}{(m^2+\kappa^2)\cos^2\theta+\sin^4\theta} \quad \text{..............(66)}.$$

Also $$\tan\delta = \frac{2\kappa\sin\theta\tan\theta}{m^2+\kappa^2-\tan^2\theta\sin^2\theta} \quad \text{..................(67)}.$$

If the optical constants κ_0 and ν_0 are known we may use equations (12) and (18) to calculate ν, κ and m, and hence $\cos 2\psi$ and $\tan\delta$ may be found.

When the difference in phase, δ, is equal to a right angle, $\tan\delta$ is infinitely large, and hence in that case
$$\sin^2\theta\tan^2\theta = m^2+\kappa^2.$$

The particular value of θ defined by this equation is called the principal angle of incidence, and corresponds to the polarizing angle in transparent media.

The problem as it generally presents itself, consists in determining the optical constants of the metal from observation of ψ and δ; for this purpose (66) and (67) are not convenient, and we must transform (65) in a different manner.

We easily deduce from that equation
$$\frac{m+i\kappa}{\sin\theta\tan\theta} = \frac{1+\tan\psi e^{i\delta}}{1-\tan\psi e^{i\delta}}$$
$$= \frac{1+\tan\psi\cos\delta + i\tan\psi\sin\delta}{1-\tan\psi\cos\delta - i\tan\psi\sin\delta},$$
and if the right side of this equation is put into the normal form, equating the real and imaginary terms gives:
$$m = \frac{\sin\theta\tan\theta\cos 2\psi}{1-\sin 2\psi\cos\delta} \quad \text{......................(68)},$$
$$\kappa = \frac{\sin\theta\tan\theta\sin 2\psi\sin\delta}{1-\sin 2\psi\cos\delta} \quad \text{..................(69)}.$$

If instead of using a compensator similar to Babinet's, the elliptic path of the disturbance of the reflected ray is analysed by a quarter wave-plate or Fresnel's rhomb, the quantities measured are the ratio of the axes of the ellipse and the inclination of these axes to the plane of incidence. Calling $\tan\Psi$ the ratio of the minor to the major axis, and γ the angle between the major axis and the plane of incidence, we obtain with the assistance of (17) (18) and (19) of Chapter I.,
$$\left.\begin{aligned} m &= \frac{\sin\theta\tan\theta\cos 2\Psi\cos 2\gamma}{1-\cos 2\Psi\sin 2\gamma} \\ \kappa &= \frac{\sin\theta\tan\theta\sin 2\Psi}{1-\cos 2\Psi\sin 2\gamma} \end{aligned}\right\} \quad \text{..............(70)}.$$

Having obtained m and κ we determine the optical constants ν_0 and κ_0 in the following manner. Equations (12) may be written

$$m^2 - \kappa^2 + \sin^2\theta = \nu_0^2 - \kappa_0^2,$$

$$m\kappa = \nu_0\kappa_0.$$

Solving these equations for ν_0 and κ_0 we obtain

$$\left.\begin{aligned} 2\nu_0^2 &= \sqrt{(m^2 - \kappa^2 + \sin^2\theta)^2 + 4m^2\kappa^2} + (m^2 - \kappa^2 + \sin^2\theta) \\ 2\kappa_0^2 &= \sqrt{(m^2 - \kappa^2 + \sin^2\theta)^2 + 4m^2\kappa^2} - (m^2 - \kappa^2 + \sin^2\theta) \end{aligned}\right\} \ldots (71).$$

Equations (68), (69) and (71) constitute the solution of our problem in the form in which Ketteler* first gave it. This form is to be preferred to the earlier one given by Cauchy, whose solution did not directly lead to the separation of the constants ν_0 and κ_0 but only to a set of equations which involved intermediate constants and variables, having no physical meaning.

In the case of metals, $m^2 + \kappa^2$ exceeds $\sin^2\theta$ sufficiently to allow us generally to neglect the square of $\sin^2\theta/(m^2 + \kappa^2)$. Under these circumstances, expressing the square root which occurs in equations (71) as a series proceeding by powers of $\sin^2\theta$, and neglecting all powers higher than the first, we find

$$\left.\begin{aligned} \nu_0 &= m\left(1 + \frac{\sin^2\theta}{2(m^2 + \kappa^2)}\right) \\ \kappa_0 &= \kappa\left(1 - \frac{\sin^2\theta}{2(m^2 + \kappa^2)}\right) \end{aligned}\right\} \ldots\ldots\ldots\ldots\ldots\ldots (72),$$

showing that as a first approximation, and especially when the angle of incidence is small, m and κ may be taken to be equal to ν_0 and κ_0.

159. The Optical Constants of Metals. We owe to Drude† the best determination of the optical constants of metals. After a careful investigation of the effects of the condition of the surface and the reflexion of surface films, results were obtained which are reproduced in Table XI. The measurements refer to sodium light.

I have added the third and fourth columns giving the values of $\kappa^2 - \nu^2$ and $\nu\kappa\cos\rho$, the two invariants of metallic refraction. The column headed θ_p gives the angle of principal incidence; the last column, the calculated reflected intensity for normal incidence.

The table shows the remarkable fact that $\nu^2 - \kappa^2$ is negative for all metals. In the older theories of refraction in which the sympathetic vibrations within the molecule were neglected, this appeared to be an anomaly, but reference to equations (38) and (39) shows that A is negative when $\omega > n$ and that $\nu^2 - \kappa^2$ may therefore also have a negative value.

* *Optik* p. 187 and *Wied. Ann.* VII. p. 119 (1879).

† *Wied. Ann.* XXXIX. p. 481.

TABLE XI.

	κ_0	ν_0	$\kappa^2 - \nu^2$	$\nu\kappa \cos \rho$	θ_p	h^2
Bismuth	3·66	1·90	9·79	6·955	77° 3′	·652
Lead	3·48	2·01	8·07	6·995	76° 42′	·621
Mercury	4·96	1·73	21·61	8·580	79° 34′	·784
Platinum	4·26	2·06	13·90	8·776	78° 30′	·701
Gold	2·82	0·37	7·81	1·043	72° 18′	·851
Antimony	4·94	3·04	15·16	15·02	80° 26′	·701
Tin	5·25	1·48	25·37	7·771	79° 57′	·825
Cadmium	5·01	1·13	23·82	5·661	79° 22′	·847
Silver	3·67	0·18	13·44	·6607	75° 42′	·953
Zinc	5·48	2·12	25·54	11·62	80° 35′	·786
Copper	2·62	0·64	6·45	1·677	71° 35′	·732
Nickel	3·32	1·79	7·82	5·943	76° 1′	·620
Cobalt	4·03	2·12	11·75	8·543	78° 5′	·675
Steel	3·40	2·41	5·75	8·194	77° 3′	·585
Aluminium	5·23	1·44	25·28	7·531	79° 55′	·827
Magnesium	4·42	0·37	19·40	1·635	77° 57′	·929

TABLE XII.

Wave-lengths in tenth-metres	4500	5000	5500	6000	6500	7000
A. Pure Metals	%	%	%	%	%	%
Platinum	55·8	58·4	61·1	64·2	66·3	70·1
Gold	36·8	47·3	74·7	85·6	88·2	92·3
Silver	90·6	91·8	92·5	93·0	93·6	94·6
Copper	48·8	53·3	59·5	83·5	89·0	90·7
Nickel	58·5	60·8	62·6	64·9	65·9	69·8
Steel (hard)	58·6	59·6	59·4	60·0	60·1	60·7
Steel (soft)	56·3	55·2	55·1	56·0	56·9	59·3
B. Speculum Metals						
Alloy of Brashear. 68·2% Cu + 31·8 Sn	61·9	63·3	64·0	64·4	65·4	68·5
Alloy of Brandes and Schünemann. 41 % Cu + 26 Ni + 24 Sn + 8 Fe + 1 Sb	49·1	49·3	48·3	47·5	49·7	54·9
Alloy of L. Mach. $66\frac{2}{3}$ % Al + $33\frac{1}{3}$ Mg	83·4	83·3	82·7	83·0	82·1	83·3
C. Glass Plates covered at the back with						
Silver	79·3	81·5	82·5	82·5	83·5	84·5
	85·7	86·6	88·2	88·1	89·1	89·6
Mercury	72·8	70·9	71·2	69·9	71·5	72·8

The question of metallic reflexion deserves further experimental treatment. Thus the observation made by Lord Rayleigh* that if a glass surface be silvered it only reflects about 83 % of the light from the glass side while it reflects over 90 % from the air side is not accounted for by the equations we have deduced.

The somewhat important question relating to the amount of light reflected at normal incidence has been investigated directly by E. Hagen and H. Rubens†. Some of their results are embodied in Table XII.

A comparison with Drude's numbers shows generally a good agreement. The alloy of Brandes and Schünemann is of practical importance owing to its permanence and resistance to deterioration when exposed to moist or impure air. Drude has also determined ν_0 for red light and found that most metals refract the red more than the yellow. The coefficient of extinction was determined directly by Rathenau‡.

Equation (39) disregarding B reduces to

$$\nu_0\kappa_0 = CV\lambda.$$

This is found to give too large a value for the coefficient of extinction in metals as already noticed by Maxwell. The disagreement is not entirely removed by introducing the term B and the theory must be recast altogether. Drude was the first to show that if conduction in metals is due to the convection of electrons the additional inertia term is sufficient to account for the partial transparency of highly conducting substances.

In the older form of the theory of conduction the total energy of the magnetic field was calculated on the assumption of a uniform distribution within each element of the conductor. But if instead of an evenly distributed electric fluid, we imagine electricity to be concentrated within the electron, the magnetic field in the immediate neighbourhood of that electron will be very much larger than the average energy in each element of volume. Hence the inertia, or the coefficient of self-induction, by whichever name the factor in question may be called, is underrated in the ordinary treatment. It may be shown that the difference affects only cases in which there is—as in the phenomena of light—a rapid variation of current. The error committed depends on the nearness of the moving electrons, and on their linear dimensions, but if their distance apart is great compared with their diameter, the additional energy per unit volume is $\frac{1}{2}\sigma i^2$ where i is the current density and σ stands for ρ/Ne^2§, N being the number of moving electrons per unit volume, and ρ the apparent mass. If each molecule

* *Collected Works*, vol. II. p. 538. † Drude's *Annalen*, vol. I. p. 352.

‡ Quoted in Winkelmann, *Encyclopædia der Wissenschaften*, vol. II. p. 838.

§ A. Schuster, *Phil. Mag.* vol. I. p. 227 (1901).

supplies one electron, which carries the conduction current, σ is of the order of magnitude 5×10^{-11} and has the dimensions of a surface.

The effect of the inertia is the same as that of an electric force $\sigma \dfrac{di}{dt}$ opposing the current. If C be the conductivity this electric force produces a current density $C\sigma \dfrac{di}{dt}$ where i only relates to the conduction current.

The equation of electric current (33) now becomes, if we denote by w' the z component of the conduction current,

$$w + C\sigma \frac{dw'}{dt} = \frac{K}{4\pi}\frac{dR}{dt} + CR + Ne\dot{\zeta},$$

also

$$w' = w - \frac{K}{4\pi}\frac{dR}{dt} - Ne\dot{\zeta}.$$

Hence if D stands as a symbol for $\dfrac{d}{dt}$

$$w + C\sigma \frac{dw}{dt} = CR + (1 + \sigma CD)\left(\frac{K}{4\pi}\frac{dR}{dt} + Ne\dot{\zeta}\right),$$

or with the help of (34), omitting the last term on the right-hand side of that equation,

$$w = CR + (1 + \sigma CD)\left(\frac{K}{4\pi}\frac{dR}{dt} + Ne\dot{\zeta}\right) - \frac{\sigma C}{4\pi}\nabla^2 R.$$

Differentiating with respect to the time and applying (34) again, we find

$$D\left\{4\pi CR + (1 + \sigma CD)\left(K\frac{dR}{dt} + 4\pi Ne\dot{\zeta}\right)\right\} = (1 + \sigma CD)\nabla^2 R \quad \text{......(73).}$$

If the motion is periodic and contains $e^{-i\omega t}$ as factor, we may substitute $-i\omega$ for D, and for $\dot{\zeta}$ we may use its equivalent (37) in terms of R. The equation then becomes

$$\nabla^2 R = -(E\omega^2 + iF\omega)R \quad \text{........................(74),}$$

where

$$E = K + NeA - \frac{4\pi\sigma C^2}{1 + \sigma^2 C^2\omega^2},$$

$$F = NeB\omega + \frac{4\pi C}{(1 + \sigma^2 C^2\omega^2)}.$$

Hence from (13)

$$\left.\begin{aligned} \nu_0^2 - \kappa_0^2 &= V^2\left(K + NAe - \frac{4\pi\sigma C^2}{1 + \sigma^2 C^2\omega^2}\right) \\ 2\nu_0\kappa_0 &= V^2\left(NBe + \frac{2C\tau}{1 + \sigma^2 C^2\omega^2}\right) \end{aligned}\right\} \quad \text{.........(75),}$$

where in the last term, τ is written for $2\pi/\omega$.

The effect of σ is therefore to diminish the product $\nu_0\kappa_0$ and hence to diminish that part of the absorption which depends on conductivity. Equations (75) are, allowing for a change of notation, identical with those obtained by Drude. If we disregard A and B, we obtain

$$\left.\begin{aligned}\nu_0^2-\kappa_0^2&=1-\frac{4\pi\sigma C^2V^2}{1+\sigma^2C^2\omega^2}\\ \nu_0\kappa_0&=\frac{CV\lambda}{1+\sigma^2C^2\omega^2}\end{aligned}\right\}\ldots\ldots\ldots\ldots\ldots\ldots(76).$$

If the numerical values of C and ω are introduced and the quantity σ is estimated, it is found that $\sigma C\omega$ is large, and equations (75) become with sufficient accuracy

$$\nu_0^2-\kappa_0^2=1-\frac{4\pi V^2}{\sigma\omega^2}+NAV^2e,$$

$$\nu_0\kappa_0=V^2\left(\frac{2\tau}{\sigma^2C\omega^2}+NBe\right).$$

As B is always po[illegible] it follows that

$$\nu_0\kappa_0>\frac{V\lambda}{C\omega^2\sigma^2}.$$

This relation allows us to calculate an upper limit for the number of electrons which take part in conduction currents, and it is found that this number in the different metals is of the same order of magnitude as the number of molecules*.

160. Reflecting powers of metals for waves of low frequency. Maxwell's theory has received an important confirmation in the work recently published by Hagen and Rubens†, on the relation between the optical and electrical qualities of metals. These investigations relate to waves of low frequency.

Neglecting σ, we may write equations (76)

$$\nu_0^2-\kappa_0^2=1,$$

$$\nu_0\kappa_0=CV\lambda.$$

As λ is supposed to be large, both ν and κ must be large and nearly equal. The second equation gives, neglecting the difference between the two quantities,

$$\kappa_0=\sqrt{CV\lambda}.$$

To test this formula for long waves, Hagen and Rubens measure the reflecting powers at normal incidence. For the intensity of the reflected light, we have obtained the expression (64), which by substitution becomes $(\kappa_0-1)/(\kappa_0+1)$ for large values of κ_0.

* Schuster, *Phil. Mag.*, Vol. VII. p. 151 (1904).

† *Ann. d. Physik*, Vol. XI. p. 873 (1903) and *Phil. Mag.* Vol. VII. p. 157 (1904).

Writing R' for the reflecting powers of a metal in per cent., $100-R'$ gives the intensity of the light which enters the metal, the intensity of incident light being 100, and the formula to be verified becomes

$$100 - R' = \frac{200}{\sqrt{C\tau}},$$

where τ is the period.

TABLE XIII.

1	2	3	4	5	6
Metals	$(100-R')$ for $\lambda=12\mu$		Conductivity at 170°	$(100-R')$ for $\lambda=25{\cdot}5\mu$ & 170°	
	Observed	Computed		Observed	Computed
Silver	1·15	1·3	39·2	1·13	1·15
Copper	1·6	1·4	32·5	1·17	1·27
Gold	2·1	1·6	27·2	1·56	1·39
Aluminium			20·4	1·97	1·60
Zinc			10·2	2·27	2·27
Cadmium			8·40	2·55	2·53
Platinum	3·5	3·5	5·98	2·82	2·96
Nickel	4·1	3·6	5·26	3·20	3·16
Tin			5·01	3·27	3·23
Steel	4·9	4·7	3·30	3·66	3·99
Mercury			0·916†	7·66	7·55
Bismuth	17·8	11·5	0·513	25·6	10·09
Rotguss*			7·05	2·70	2·73
Manganin			2·37	4·63	4·69
Constantin	6·0	7·4	2·04	5·20	5·05
Patent Nickel P.	5·7	5·4	3·69	4·05	3·77
Patent Nickel M.	7·0	6·2	2·86	4·45	4·28
Rosse's Alloy	7·1	7·3			
Brandes and Schunemann's Alloy	9·1	8·6			

In Table XIII., columns 2, 3, the observed and computed values of $100-R'$ are tabulated for $\lambda = 12\mu$ ($\mu = 10^{-4}$ cms.). For larger values of λ, R' approaches 100 % asymptotically with increasing wave-lengths, and the difficulty of experimentally determining $100-R'$ increases accordingly. Hagen and Rubens therefore measured the emissive power instead of the reflecting power of the metals.

* "Rotguss" contains 85·7 Cu + 7·2 Zn + 6·4 Sn.

† At 100°.

From a comparison of the radiation sent out by a metal with that sent out by a black body, the reflecting power may be directly deduced. In Table XIII., column 4 gives the conductivity at 170° C. calculated from the known conductivities at 18° and the temperature coefficient. Columns 5 and 6 give the observed and computed emissive powers at 170°. It will be noticed that the agreement is excellent in all cases except Aluminium which shows a considerable deviation, and Bismuth which forms a complete exception to the law.

Professors Hagen and Rubens have also directly verified the fact that the quantity $100 - R'$ indicates with increasing temperature a change corresponding to the change of electrical resistance, and they further point out the remarkable fact that it would be possible to undertake absolute determinations of electrical resistance solely by the aid of measurements on radiation. The agreement of Maxwell's theory in its simple original form with the result of the experiments just described, proves that for wave-lengths as great as 12μ, the free periods of vibration of the molecules do not affect the optical constants of metals.

161. Connexion between refractive index and density. All investigations discussed in this Chapter which base the optical properties of a medium solely on the responsive vibrations of the atoms and molecules appear to lead to the conclusion that $\mu^2 - 1$ is proportional to the density. Changes of temperature and pressure should only affect the refractive index in so far as the density is changed. Experiments do not confirm this conclusion except in the case of gases. A good example of the failure of the formula is furnished by water near the freezing point. The variations in density, which have a well-defined maximum at 4° C., have no counterpart in corresponding changes of the refractive index. We must conclude that the theory is incomplete and leaves essential factors out of account. If we examine the assumptions tacitly made, we can discover at any rate two which may contribute to the discord between theory and experiment. We have assumed in the first place that the free vibrations of the electrons within the molecule are unaffected by changes of temperature and pressure. It is likely, on the contrary, that any changes in chemical constitution or of molecular grouping materially affect the result. An even more important cause of the discordance lies in the assumption that the molecules act independently of each other. If—as in the case of solids—they are more or less regularly spaced, the formulae need correction. Two authors of similar name, H. A. Lorentz* of Leyden, and L. Lorenz† of Kopenhagen, have almost simultaneously published

* *Wied. Ann.* IX. p. 641. (1880.)
† *Wied. Ann.* XI. p. 70. (1880.)

investigations which take account of the mutual effects of the molecules and lead to the conclusion that it is $(\mu^2-1)/(\mu^2+2)D$ which is constant. (D denotes the density.) The formula has been confirmed in many cases by experiment. It is remarkable that the formula of Lorentz and Lorenz correctly gives for a number of substances the alteration of the refractive index during the change from the liquid to the gaseous state. The effect of regular molecular spacing on physical properties has also been investigated by the late Lord Rayleigh*.

An attempt has been made by Gladstone and Dale to connect the refractive indices of a compound with that of the elements composing it. They concluded that if any compound of molecular weight ω and density D contains $n_1 n_2 \ldots$ atoms of various kinds, the refractive index of the compound could be calculated approximately from the formula

$$\frac{(\mu-1)\,\omega}{D} = n_1 v_1 + n_2 v_2 + n_3 v_3\, r \ldots,$$

when v_1, v_2, v_3 are constant for each element.

162. Historical. Augustin Louis Cauchy, whose work has already been referred to at the end of Chapter x., published some important researches in wave propagation and first obtained formulae giving the constants of elliptic polarization of light on reflexion from metallic surfaces. These he published however without proof.

Jules Céléstin Jamin (born May 30, 1818, in the Department of the Ardennes, died February 12, 1886, at Paris) was the pioneer in the experimental investigation of metallic reflexion, and showed that Cauchy's equations represented the facts with sufficient accuracy. Eisenlohr supplied the analytical proof of Cauchy's formulae and showed that the proper interpretation of Jamin's measurements leads to the conclusion that for silver the refractive index is smaller than one. This result, which did not seem at that time to be reconcilable with the stability of the medium inside the metal, received support from Quincke's experiments which proved an *acceleration* of phase when light passed through thin metallic films. The matter was finally settled by A. Kundt† (born Nov. 18, 1839 at Schwerin, died May 21, 1894 near Lübeck, Professor of Physics in the University of Berlin) who succeeded in making thin prisms of metals and thus could demonstrate directly that in metals light was propagated more quickly than in vacuo. The apparent anomaly of this result received its explanation when the refraction of absorbing media generally was more carefully studied.

In 1862 Le Roux having filled a hollow prism with the vapour of iodine, noticed that while it absorbed the central parts of the spectrum,

* *Collected Works*, Vol. IV. p. 19.

† *Wied. Ann.* XXXIV. p. 469. (1888.)

it transmitted the red and violet ends, refracting however the red end more than the violet. This phenomenon he called anomalous dispersion. Eight years later, Christiansen noticed the same phenomenon in the case of a solution of fuchsin. The matter then attracted considerable attention, and A. Kundt especially improved the experimental methods. Including a great many colouring matters in his investigations he was able to formulate the general laws which regulate the influence of absorption on refraction. In the meantime, Sellmeyer had published his theoretical investigation, which is now generally recognized to be correct in principle. It only remains to allude to the work of Ketteler, who more than any one else has shown, both by experiment and by mathematical calculation, that all refraction is of one kind, and that even in the case of apparently transparent media like water, it is necessary to take account of the effects of the free vibrations of the molecules both in the infra-red and ultra-violet.

The recent development of the subject has already been sufficiently treated.

CHAPTER XII.

ROTATORY EFFECTS.

163. Photo-gyration. In all cases hitherto considered the transmission of a luminous disturbance has been such that a plane polarized wave was propagated with its plane of polarization remaining parallel to itself. But there are media in which the wave, though remaining plane polarized, shows a continuous rotation of the plane of polarization as it proceeds. If plane polarized light be made to traverse, for instance, a tube filled with a sugar solution, and the emergent light be examined, it is observed that the plane of polarization has been turned through an angle which depends on the concentration of the solution and is proportional to the length of the tube.

The direction of rotation may be right-handed or left-handed. It is said to be right-handed when it is in the direction of the rotation of the hands of a watch, looked at from the side towards which the light travels.

Substances which possess this property are often called "optically active," an expression which is not very descriptive and possibly misleading, as the word "activity" has been applied to several different properties. We shall find that the distinctive feature of the rotational property is the different velocity of propagation of circularly polarized light according as it is right-handed or left-handed. We may therefore appropriately call substances "dextrogyric" or "laevogyric" according as they turn the plane of polarization to the right or to the left (γῦρος, a circle). A substance is simply called "photogyric" if it acts in its isotropic state, but "crystallogyric" if, like quartz, the property is connected with its crystalline nature. Finally all substances turn the plane of polarization when they are traversed by light in the direction of a magnetic field. They become therefore "magneto-gyric." If a special word be required to express the general property not applied to any particular case, we shall use the expression "allogyric" (ἄλλος, different), while substances which are inactive are "isogyric."

The allogyric property implies some asymmetrical structure, and in the case of solutions, the want of symmetry must be in the structure of the molecule itself. Van 't Hoff and le Bel have indeed drawn important conclusions as to the arrangement of the atoms in the molecule of allogyric substances.

Quartz is the most conspicuous example of a crystallogyric body. If a plate a few millimetres thick be cut out of a crystal of quartz perpendicularly to the axis, and this plate be inserted between crossed Nicols, the luminosity of the field is seen to be restored. If the original light was white, the transmitted light is coloured. The explanation of the effect presents no difficulty on the assumption of a rotation of a plane of polarization which is different for different wave-lengths. There is no rotation of the plane of polarization if the wave-front is parallel to the axis. Some specimens of quartz show a right-handed rotatory effect while others are left-handed. It is found that generally the direction of the rotation may be detected by a close examination of the crystal, there being certain small asymmetrical planes at the corners between the hexagonal prism and pyramid, the position of which is different for the two types of crystals. In all substances hitherto discovered, which are allogyric, the angle of rotation, per unit length of substance traversed, increases with the refrangibility and is approximately proportional to the inverse square of the wave-length.

There is a marked distinction between the magnetogyric and other allogyric effects. In the case of substances which possess the rotatory property in their natural state, the rotation for rays travelling opposite ways is in the same direction when looked at from the same position *relative to the direction in which the light travels.* Thus if A and B are two ends of a tube containing a solution of sugar and light is sent through the tube from A to B, an observer looking at B towards the light will observe a right-handed rotation. If now the light be sent from B to A and the observer looks at A, the rotation observed by him is still right-handed. If there were a mirror at B, and the ray after traversing the tube from A to B were reflected back towards A, the plane of polarization at emergence would be parallel to the direction it had on first entering the tube at A. This we should indeed expect by the principle of reversibility (Art. 25). In the case of magnetogyration on the contrary the direction of rotation is different as seen by the observer according as the light travels with or against a line of force, but it is the same when looked at from the same position *relative to the direction of the magnetic field.* Consequently if light travels from A to B, and is reflected back at B, the angle through which the plane of polarization is rotated is increased and finally doubled during the passage backwards. The principle of reversibility holds in this case also, but we must reverse the direction of the magnetic field as well as the direction of the ray.

164. Analytical representation of the rotation of the plane of polarization. Consider plane waves travelling in the direction of x, with a uniformly rotating direction of vibration. As each wave-front reaches a given position, the direction of vibration is a definite one, and the angle which that direction forms with one fixed in space is therefore a function of x only. If it be a linear function of x, the plane of polarization rotates through an angle which is proportional to the distance traversed. Let η and ζ be the projections of the displacement, and put

$$\left.\begin{aligned}\eta &= 2\cos rx \cos(lx-\omega t)\\ \zeta &= 2\sin rx \cos(lx-\omega t)\end{aligned}\right\} \quad\ldots\ldots\ldots\ldots\ldots\ldots\ldots(1).$$

The equations satisfy the condition laid down for the direction of vibration, for if δ be the angle between it and the axis of z

$$\begin{aligned}\tan\delta &= \frac{\eta}{\zeta}\\ &= \tan rx,\end{aligned}$$

from which it follows that δ is a linear function of x, and that r measures the angle of rotation per unit length of path. We call the quantity r the "gyric coefficient." Equations (1) also satisfy the conditions of ordinary wave propagation, as the displacements may be expressed as a sum of terms, each of which has the form $f(x-vt)$. To show this we need only transform the products of the circular functions in a well-known manner.

Writing

$$\left.\begin{aligned}\eta_1 &= \cos(l_1x-\omega t); \quad \eta_2 = \cos(l_2x-\omega t)\\ \zeta_1 &= \sin(l_1x-\omega t); \quad \zeta_2 = -\sin(l_2x-\omega t)\end{aligned}\right\} \quad\ldots\ldots\ldots(2),$$

we find that (1) becomes identical with

$$\eta = \eta_1 + \eta_2; \qquad \zeta = \zeta_1 + \zeta_2,$$

provided that

$$r = \tfrac{1}{2}(l_1 - l_2); \quad l = \tfrac{1}{2}(l_1 + l_2),$$

or if r and l be given

$$l_1 = l + r; \quad l_2 = l - r.$$

The disturbance is now expressed in terms of four parts, each of which is of the homogeneous type, but while the periodic time for each of these four waves is the same, the wave-lengths are in groups of two: $2\pi/l_1$ and $2\pi/l_2$ respectively. The displacements η_1 and ζ_1 form together a right-handed circularly polarized ray, propagated with velocity $v_r = \omega/(l+r)$, while the displacements η_2 and ζ_2 combine to form a left-handed circularly polarized ray propagated with velocity $v_l = \omega/(l-r)$. The gyric coefficient may be deduced from v_r and v_l by means of

$$r = \frac{\omega}{2}\left(\frac{1}{v_r} - \frac{1}{v_l}\right) \qquad (3).$$

The important conclusion that a wave travelling with a uniform rotation of its planes of polarization is equivalent analytically to the superposition of two circularly polarized rays of opposite directions and propagated with different velocities is due to Fresnel. A simple geometrical illustration may be given. If two points P and Q are imagined to revolve in opposite directions with uniform and identical velocities round the circumference of a circle (Fig. 174), they will cross at two opposite ends A and B of a diameter, and their combined motion is equivalent to a simple periodic motion along AB as diameter. The two points may be considered to represent the displacements of two waves polarized circularly in opposite directions, having for their resultant a plane polarized wave. If the two circularly polarized waves are transmitted with different velocities, there is, as the waves proceed, a gradual retardation of one circular motion relative to the other, so that the crossing points gradually shift to one side or the other. The combined motion always remains a simple periodic motion along a diameter, but that diameter rotates uniformly as we proceed along the wave normal. If A_1, B_1 are the crossing points in a wave-front which is at unit distance from that originally considered, AOA_1 represents the angle through which the plane of polarization is turned in unit length of path.

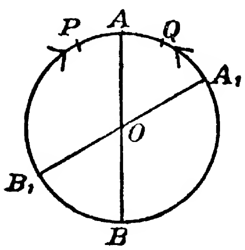


Fig. 174.

165. Isotropic substances. There is no satisfactory representation of the mechanism by means of which an asymmetrical molecular structure turns the plane of polarization, but we may easily extend our former equation so as to include rotatory effects. In the equation (Art. 152, Chapter XI.):

$$\ddot{\zeta} + n^2\zeta = \frac{eR}{\rho} \qquad (4),$$

it was assumed that the electron suffers no constraint in its motion; but if forces act which depend on the displacements of other electrons, the resultant force may involve not only the three components P, Q, R of electric force, but also their nine differential coefficients with respect to the three independent space variables. Considering small motions only, we need only take linear terms into account. The complicated general equation which would result from the substitution on the right-hand side of (4) of twelve linear terms is much simplified by the restriction of our investigation to isotropic substances.

In such substances a luminous wave is affected equally in whatever direction it passes, and the resultant differential equation must therefore be independent of the direction of the coordinate axes. If for instance we turn the system of axes through 180° round the axis of z, the simultaneous reversal of the signs of P, Q, x and y must leave the equations unaltered. This consideration shows that there are no terms involving Q, R, $\frac{dP}{dz}$, $\frac{dQ}{dz}$, $\frac{dR}{dx}$ and $\frac{dR}{dy}$ because all these terms if existing would reverse their sign by the supposed change in the coordinate axes. Similarly if we rotate the axes through 180° round the axis of x, the left-hand side of (4) changes sign, hence the general term to be substituted on the right must also reverse its sign. This excludes the terms depending on $\frac{dP}{dx}$, $\frac{dQ}{dy}$, $\frac{dR}{dz}$. The only remaining differential coefficients are $\frac{dP}{dy}$ and $\frac{dQ}{dx}$, and these must occur in the combination $\frac{dP}{dy}-\frac{dQ}{dx}$, as may be seen by turning the system through 90° round the axis of z and introducing the condition that the equation remains unaltered. We may therefore write the resulting differential equation:

$$\ddot{\zeta}+n^2\zeta=\frac{e}{\rho}\left[R+s\left(\frac{dP}{dy}-\frac{dQ}{dx}\right)\right] \quad\text{..................(5).}$$

Similarly

$$\ddot{\eta}+n^2\eta=\frac{e}{\rho}\left[Q+s\left(\frac{dR}{dx}-\frac{dP}{dz}\right)\right] \quad\text{..................(6).}$$

Confining ourselves to insulators, equation (35) of Chapter XI. is

$$K\frac{d^2R}{dt^2}=\nabla^2R-4\pi Ne\ddot{\zeta} \quad\text{..........................(7).}$$

If the displacements are proportional to $e^{-i\omega t}$, so that in (5) we may write $-\ddot{\zeta}/\omega^2$ for ζ, we obtain by substitution in (7)

Similarly

$$\left.\begin{aligned} K\frac{d^2R}{dt^2}&=\nabla^2R-\omega^2m\left[R+s\left(\frac{dP}{dy}-\frac{dQ}{dx}\right)\right]\\ K\frac{d^2Q}{dt^2}&=\nabla^2Q-\omega^2m\left[Q+s\left(\frac{dR}{dx}-\frac{dP}{dz}\right)\right]\end{aligned}\right\} \quad\text{...........(8),}$$

where

$$m=\frac{4\pi Ne^2}{\rho\,(\omega^2-n^2)}.$$

If we consider plane waves parallel to yz so that the electric forces are independent of y and z, equations (8) may be written

$$\left.\begin{aligned}\frac{d^2R}{dx^2} &= -\left(M_1R - M_2\frac{dQ}{dx}\right)\\ \frac{d^2Q}{dx^2} &= -\left(M_1Q + M_2\frac{dR}{dx}\right)\end{aligned}\right\}\quad\text{.................. (9),}$$

where $$M_1 = \omega^2(K+m), \quad M_2 = \omega^2 ms.$$

From equations (9) we derive:

$$\frac{d^2}{dx^2}(Q+iR) = -\left[M_1 - iM_2\frac{d}{dx}\right](Q+iR) \text{ .... (10).}$$

This has for a particular solution

$$Q + iR = e^{i(l_1x-\omega t)} \text{ .... (11),}$$

provided that

$$l_1^2 = M_1 + M_2l_1 \text{ .... (12).}$$

Reversing the sign of i in (10) and assuming a solution

$$Q - iR = e^{i(l_2x-\omega t)} \text{ .... (13),}$$

we find the equation of condition

$$l_2^2 = M_1 - M_2l_2 \text{ .... (14).}$$

The positive roots of (12) and (14) which alone need be considered are

$$\left.\begin{aligned}2l_1 &= M_2 + \sqrt{M_2^2 + 4M_1}\\ 2l_2 &= -M_2 + \sqrt{M_2^2+4M_1}\end{aligned}\right\}\text{ .... (15).}$$

Separating and retaining only the real parts in the solutions (11) and (13), it is seen by comparison with (2) that (11) represents a right-handed circular polarization, while (12) represents a left-handed circular polarization. The two waves are propagated with velocities ω/l_1 and ω/l_2 respectively. The superposition of both solutions represents a plane polarized wave, the plane of polarization rotating per unit length of optical path through an angle r which is obtained from (3):

$$\begin{aligned}r &= \tfrac{1}{2}(l_1 - l_2) = \tfrac{1}{2}M_2\\ &= \frac{2\pi Ne^2s}{\rho}\,\frac{\omega^2}{\omega^2-n^2}\end{aligned}\text{ .... (16).}$$

The above investigation shows that the only terms depending on the first differential coefficients of the electric forces which can be added to the general equations of light and are consistent with isotropy indicate a turn of the plane of polarization. This does not of course furnish an explanation of the rotatory effect, which would require a knowledge of the physical cause to which the terms are due. We may however take one step forward towards an explanation by considering that the terms in equation (5) which have been added represent a torsional electric force having the axis of ζ as axis. The equations mean therefore that a displacement of the electron in the z direction may be produced not only by a force acting in that direction, but also by a couple acting round it. A rifle bullet lying in its rifle barrel would be displaced in a similar manner along the barrel both by a pulling and twisting force. But if we take the dimensions of a single electron to be very small, we exclude the possibility of a constraint which would enable a couple to cause a motion in one direction. We must in that case draw the conclusion that the vibrations of the electron which give rise to the rotatory effect are motions of systems of electrons united together by certain forces which are such that a couple of electric forces produces a displacement of the positive electrons in one direction or of the negative electrons in the opposite direction along the axis of the couple. In view of the fact that a single electron cannot be acted on by a torsional force, it would have been more appropriate to base our investigation on equation (44) Chapter XI. The generalized force Ψ_1 would in the present problem depend not only on R but on $\left(\frac{dP}{dy} - \frac{dQ}{dx}\right)$, and if the investigation in the second part of Art. 152 is modified by the addition of appropriate terms, the result arrived at would, for a single variable, remain the same as that represented by (16).

166. Allogyric double refraction. Equations (2) show that the analytical representation of plane polarized waves travelling through an optically active medium necessarily involves two different wave velocities. In any question concerning the refraction and reflexion of light, we may take all four displacements represented by (2) separately and apply the formula obtained for homogeneous disturbances. It is clear that the wave on emergence must be split into two separate waves which are circularly polarized in opposite directions This double refraction, due to the rotational effect, is verified by experiment and has some practical importance. Quartz, as has already been mentioned, turns the plane of polarization of waves travelling parallel to the optic axis, and in consequence, a ray travelling along the optic axis is doubly refracted at emergence. Quartz is very useful in optical investigations on account of its transparence to ultra-violet rays, and it is a serious drawback that it is impossible to avoid double refraction in a prism made of that substance. The difficulty is overcome by combining two prisms made of two specimens, one of which has a right-handed and the other a left-handed rotatory power These two prisms ABC and A_1BC (Fig. 175) are right-angled at C and have their optic axes parallel to AA_1. They are joined together along BC, and if a ray traverses such a prism at minimum deviation its direction inside the prism is parallel to the axis. A ray polarized either in the principal plane or at right angles to it, is divided into two rays circularly polarized in opposite directions. The same is true therefore of an unpolarized ray. Of these circularly polarized rays one gains over the other while traversing the first prism, and loses equally while passing through the second prism. The combined optical distance is therefore the same for both components, and there is only a single refraction at emergence.

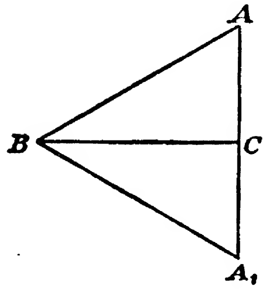


Fig. 175.

167. Crystalline media. A plane wave travelling through quartz splits up into two plane polarized waves if the wave travels at right angles to the axis, and into two circularly polarized waves if it travels parallel to the axis. In the case of waves travelling obliquely to the axis we may therefore surmise that the two waves are elliptically polarized, the ellipse becoming more and more eccentric as the wave becomes less inclined to the axis. This conclusion is verified by experiment. The elements of the ellipse have been made the subject of calculation by Sir George Airy*. A very clear account of this subject is given by Mascart†. It was believed for some time that crystallogyric properties were confined to uniaxal crystals, but Pocklington‡ observed effects in the case of cane sugar. Subsequently Dufet§ found other biaxal crystals showing a turn of the plane of polarization depending on the orientation of the crystals.

168. Rotatory dispersion. The rotation per unit length according to (16) is

$$r = \frac{\beta\omega^2}{n^2 - \omega^2},$$

on the supposition that we need only consider one period $2\pi/n$ of the free vibration. In this expression β is a constant which may either be positive or negative according to the sign of s. If the free period is very short compared with the range of visible periods, we may neglect ω in comparison with n, and the rotation is in that case proportional to ω^2, *i.e.* inversely proportional to the square of the wave-length. This law holds approximately for most substances which have been examined. In general we have to consider several free periods, so that we must write

$$r = \Sigma \frac{\beta_m\omega^2}{n_m^2 - \omega^2} \qquad (17),$$

* *Camb. Phil. Trans.*, Vol. IV. Part 1 (1831).
† *Optique*, Vol. II. p. 513.
‡ *Phil. Mag.* II. p. 361 (1901).
§ *Bull. Soc. Min.* XXVII. p. 156.

the summation having to be carried out for the different values of m. If the free periods lie in the ultra-violet so that all values of n_m are larger than ω we may expand the function in powers of ω and obtain after rearranging terms

$$r = r_1\omega^2 + r_2\omega^4 + r_3\omega^6 \text{ ....} \quad (18),$$

where r_1, r_2, r_3, are quantities depending on the values of β_m and n_m The rotatory properties of quartz have been investigated over a very wide range. It is found that the effects may be explained by assuming two ultra-violet free periods, one of which may be made to coincide with the ultra-violet period, which has been deduced from the general dispersion effects of quartz (Art. 153), the other being very short*. The infra-red periods necessary for the explanation of refraction do not seem to produce any rotatory effects.

169. Isochromatic and achromatic lines. The appearance of photogyric crystals in the polariscope is materially affected by their rotatory effect. The calculation of the isochromatic and achromatic lines has been carried out by Sir George Airy. A full account is given in Mascart's *Optics*†. The simplest case is that of a plate cut at right angles to the axis examined with crossed polarizer and analyser. Apart from the rotatory effect, the appearance should be that of Fig. 1, Plate II. Now owing to this rotatory effect the vibration which enters near the centre parallel to the principal plane of the polarizer leaves it inclined at an angle to that direction and is not therefore completely blocked out by the analyser. The result is that there are no achromatic lines near the centre. The general appearance is that of the figure, omitting the dark cross within the first dark ring.

170. The Zeeman effect. Before discussing the theory of photogyric effects, which a magnetic field impresses on a wave of light passing through it, we may give a short account of the modifications of the luminous radiations observed when the source of light is subjected to strong magnetic forces. It was discovered by Zeeman in 1896 that a sodium flame placed in a magnetic field showed a widening of the two yellow lines, and at the suggestion of H. A. Lorentz, who at once foresaw the right explanation, further experiments were made to test the polarization of the emitted radiations which confirmed Lorentz's theory. In the case of spectroscopic lines, which show the simplest type of magnetic effect, it is found that if the light is examined axially, *i.e.* parallel to the lines of force, each line splits into two, which are circularly polarized in opposite directions. Looked at

* Drude, *Optik*, p. 381.

† *Optique*, Vol. II. p. 314.

equatorially, each line is divided into three components, the centre one being polarized in an equatorial plane, and the two others in a plane passing through the lines of force.

If we look upon the radiations as being due to the vibrations of an electron these observations admit of a simple explanation. Consider, first, light sent out in the axial direction. Each rectilinear vibration may be supposed to be made up of two opposite circular vibrations, the orbits lying in the equatorial plane. Let the light which reaches the observer travel through the flame in the direction of the lines of force, *i.e.* from the north to the south magnetic pole. A positive electron performing an anti-clockwise rotation, *i.e.* a positive rotation round a line of force, will under these circumstances be acted on by a force Hev, tending to increase the diameter of the circle in which it revolves (H = intensity of magnetic field, v = linear speed of electron, e = charge of electron). If in the absence of the magnetic field the acceleration is n^2d, where d is the displacement, the force when the magnet is excited is $\rho n^2 d - Hev$, ρ being the mass. But if $2\pi/n_1$ be now the periodic time, then as v is the velocity and therefore $2\pi d/v$ the time in which the circle is described, it follows that $v = n_1 d$ and that the acceleration of the particle is $n_1^2 d$. Hence

$$\rho n^2 d - Hev = \rho n_1^2 d,$$

and substituting $n_1 d$ for v

$$\rho(n^2 - n_1^2) = Hen_1.$$

As the difference between n and n_1 is small, we may now write with sufficient accuracy

$$n - n_1 = \frac{He}{2\rho}.$$

Finally, introducing the frequencies N and N_1 in place of n and n_1 we obtain

$$N - N_1 = \frac{He}{4\pi\rho},$$

or if we write

$$z = \frac{e}{4\pi\rho} \quad \text{..............................(20)},$$

$$N - N_1 = zH \quad \text{..............................(21)}.$$

The coefficient z may conveniently be called the Zeeman coefficient. We conclude that a rectilinear simply periodic motion is divided into two circular motions, the longer period showing anti-clockwise rotation if e is positive. Zeeman observed that the less refrangible component rotates clockwise, and the more refrangible one anti-clockwise, if the field is in the specified direction, and it follows that if our theory is correct it is the negative electron that gives rise to all vibrations for which this is the case.

Looked at equatorially, the two circular orbits appear in projection as vibrations which are rectilinear, and at right angles to the lines of force. The vibrations which take place along the lines of force are not, in the simple theory here considered, affected by the magnetic field. They constitute therefore plane polarized light with unchanged period. These vibrations have no component however which can be transmitted in the axial direction. Looked at equatorially we should expect therefore to see each line divided into three, the external components of the triplet having the same period as the circular vibrations observed in the axial direction. The agreement of the appearance, reasoned out in this fashion, with the observed facts constitutes a direct proof that the direction of vibration is at right angles to the plane of polarization, if we identify variations of electric force with direction of vibration. Although there is much indirect evidence in favour of this view, such a convincing demonstration as that afforded by the Zeeman effect is very satisfactory.

From equations (20) and (21) we may devise a value of e/ρ by measuring $N - N_1$ in terms of H. In an extended research Runge and Paschen* have determined the Zeeman coefficient for a number of spectroscopic lines for which the separation takes place in accordance with the theoretical law. The devised values of e/ρ lie between $3{\cdot}4 \times 10^7$ and 10^7, while most reliable direct measurements of that constant by Millikan give $1{\cdot}77 \times 10^7$. We conclude that at any rate as regards order of magnitude the theory is in accord with experiment.

Our calculation has tacitly assumed that the vibrating electron is free from constraint and acts as an independent unit with three degrees of freedom. If we drop this assumption we are led to more complicated magnetic effects, and indeed the majority of spectroscopic lines do not show a more complex subdivision than is indicated by the theory in its simple form.

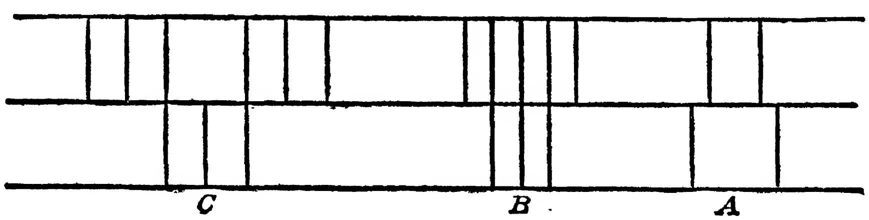


Fig. 176.

Fig. 176 shows three remarkable cases†. In this figure the components vibrating perpendicularly are drawn above those which are parallel to the magnetic field. The peculiarity of the type marked A is that the vibrations parallel to the field are more affected than those at right angles to it. According to Berndt the green line of

* *Berl. Abh.* (Anhang) 1902. *Sitzungsber. d. Berl. Ak.* XIX. p. 380 (1902) and XXXII. p. 720 (1902).

† *C. R.* CXXVII. p. 18 (1898).

Helium is divided in accordance with this type A, but the experiment is difficult in the case of permanent gases and further measurements are much needed. It will be noticed that in some cases the same component appears, whether the direction of vibration is at right angles to the field or parallel to it. It would be interesting to notice whether in such cases the light is really elliptically polarized, as it should be if the coincidence were absolute.

The simplest form of the theory assumes that the vibrations parallel to the lines of force preserve their period, but there are important cases in which these also change and two vibrations, one of larger and one of shorter period, take the place of the original one. In some cases the original period is maintained as well; in other cases it completely disappears. Such a phenomenon shows that the vibration along the line of force is not free, but is accompanied by changes in directions at right angles to itself, and that the magnetogyric effect of the accompanying changes reacts on the original vibration.

H. A. Lorentz* in a general theoretical discussion shows that if a spectroscopic line divides into n components, there must be n degrees of freedom in the system which in the absence of the magnetic field are coincident.

There is a well-defined relationship between the type of magnetic separation and the grouping of spectroscopic lines which has been more specially investigated by Runge and Paschen†.

It is a significant fact that no Zeeman effect has yet been observed in the case of spectra of fluted bands such as those of carbon and nitrogen‡. The magnetogyric properties of gases giving by absorption spectra of fluted bands render it very possible that such effects exist but have not been detected owing to their smallness. A slight increase in power may bring them to light.

171. Photo-gyration in the magnetic field. If an electron attracted to a fixed centre with a force varying as the distance moves in a magnetic field, its equations of motion are

$$\left.\begin{aligned}\frac{d^2\xi}{dt^2}+n^2\xi&=\frac{e}{\rho}\left[P+\left(H_3\frac{d\eta}{dt}-H_2\frac{d\zeta}{dt}\right)\right]\\ \frac{d^2\eta}{dt^2}+n^2\eta&=\frac{e}{\rho}\left[Q+\left(H_1\frac{d\zeta}{dt}-H_3\frac{d\xi}{dt}\right)\right]\\ \frac{d^2\zeta}{dt^2}+n^2\zeta&=\frac{e}{\rho}\left[R+\left(H_2\frac{d\xi}{dt}-H_1\frac{d\eta}{dt}\right)\right]\end{aligned}\right\}\quad\text{...........(22)},$$

where H_1, H_2, and H_3 are the components of magnetic induction due

* *Rapports présentes au congrès international de Physique de* 1900. Vol. III. p. 1.

† *Sitzungsber. d. Berl. Ak.* XIX. p. 380, and XXXIII. p. 702.

‡ *C. R.* CXXVII. p. 18 (1898).

to the external field. The right-hand side of the equations which express the components of electromagnetic force may easily be proved from the consideration that the force is at right angles both to the direction of the field and the direction of motion*.

If we take the magnetic field to be of uniform strength H, the lines of force being parallel to the axis of x, the above equations may be written more simply

$$\left.\begin{aligned}\ddot{\xi}+n^2\xi&=\frac{e}{\rho}P\\ \ddot{\eta}+n^2\eta&=\frac{e}{\rho}(Q+H\dot{\zeta})\\ \ddot{\zeta}+n^2\zeta&=\frac{e}{\rho}(R-H\dot{\eta})\end{aligned}\right\}\quad\text{......................(23).}$$

These equations together with

$$\left.\begin{aligned}K\ddot{P}&=\nabla^2P-4\pi Ne\ddot{\xi}\\ K\ddot{Q}&=\nabla^2Q-4\pi Ne\ddot{\eta}\\ K\ddot{R}&=\nabla^2R-4\pi Ne\ddot{\zeta}\end{aligned}\right\}\quad\text{......................(24)}$$

determine the problem.

If a plane wave be propagated parallel to a line of force, P and ξ vanish, and by elimination of η and ζ between (23) and (24) we may obtain two equations which only contain P and Q. For the sake of simplicity, we shall confine ourselves to the simple periodic motion.

Writing $-i\omega$ for d/dt and $-\omega^2$ for d^2/dt^2 and introducing symbols σ and Π defined by

$$\begin{aligned}\sigma_r&=\eta+i\zeta; &\quad \Pi_r&=Q+iR,\\ \sigma_l&=\eta-i\zeta; &\quad \Pi_l&=Q-iR,\end{aligned}$$

we obtain from (23) and (24)

$$\left.\begin{aligned}\sigma_r\{(n^2-\omega^2)\rho+He\omega\}&=e\Pi_r\\ \sigma_l\{(n^2-\omega^2)\rho-He\omega\}&=e\Pi_l\end{aligned}\right\}\quad\text{..................(25),}$$

$$\left.\begin{aligned}K\ddot{\Pi}_r&=\nabla^2\Pi_r-4\pi Ne\ddot{\sigma}_r\\ K\ddot{\Pi}_l&=\nabla^2\Pi_l-4\pi Ne\ddot{\sigma}_l\end{aligned}\right\}\quad\text{......................(26).}$$

$\ddot{\sigma}_r$ and $\ddot{\sigma}_l$ may now be eliminated, and we derive thus from (25) and (26),

$$\left(K+\frac{4\pi Ne^2}{(n^2-\omega^2)\rho+He\omega}\right)\frac{d^2\Pi_r}{dt^2}=\frac{d^2\Pi_r}{dx^2},$$

$$\left(K+\frac{4\pi Ne^2}{(n^2-\omega^2)\rho-He\omega}\right)\frac{d^2\Pi_l}{dt^2}=\frac{d^2\Pi_l}{dx^2}.$$

This gives for v_r the velocity of right-handed circularly polarized

* Maxwell, *Electricity and Magnetism*, Vol. II. p. 227.

light, and for v_l the velocity of left-handed circularly polarized light,

$$\left.\begin{aligned}\frac{1}{v_r^2} &= K + \frac{4\pi Ne^2}{(n^2-\omega^2)\rho + He\omega}\\ \frac{1}{v_l^2} &= K + \frac{4\pi Ne^2}{(n^2-\omega^2)\rho - He\omega}\end{aligned}\right\} \quad \ldots\ldots\ldots\ldots\ldots\ldots(27),$$

and

$$\frac{1}{v_l^2} - \frac{1}{v_r^2} = \frac{8\pi e^3 NH\omega}{(n^2-\omega^2)^2\rho^2 - H^2e^2\omega^2} \quad \ldots\ldots\ldots\ldots\ldots(28).$$

In actual observation, it is difficult to apply a field much greater than 30,000 units. Assuming for e/ρ the value $1{\cdot}6 \times 10^7$, we find for He/ρ, $4{\cdot}8 \times 10^{11}$. Also ω for green light is $1{\cdot}3 \times 10^{15}$. If n and ω differ by not less than the two-hundredth part of their period, so that $n - \omega = 6 \times 10^{12}$ or more, the second term in the denominator of (28) is equal to less than the six-hundredth part of the first and may be neglected.

We may write under these circumstances, if v represents the velocity of light in the absence of a magnetic field,

$$\frac{1}{v_l^2} - \frac{1}{v_r^2} = \frac{2}{v}\left(\frac{1}{v_l} - \frac{1}{v_r}\right) = \frac{8\pi e^3 NH\omega}{(n^2-\omega^2)^2\rho^2}.$$

The gyric coefficient (r) is equal to $\frac{1}{2}\omega\left(\frac{1}{v_r} - \frac{1}{v_l}\right)$ and hence

$$\begin{aligned} r &= -\frac{2\pi e^3 NH\omega^2 v}{(n^2-\omega^2)^2\rho^2}\\ &= -\frac{\beta\omega^2 vH}{(n^2-\omega^2)^2} = -\frac{\beta\omega^2 VH}{\mu(n^2-\omega^2)^2} \quad \ldots\ldots\ldots\ldots(29),\end{aligned}$$

where β is an appropriate coefficient, and μ the refractive index.

If the free periods are much more rapid than those to which the observations apply, ω in the denominator of (29) may be neglected and r is approximately proportional to ω^2, which agrees with the experimental facts. If the electron, the motion of which has been considered, is positive, the rotation has the opposite sign to H. As we take a clockwise rotation as negative, this means that the rotation is right-handed when the light travels in the direction of a line of force (from North to South). If the vibrating electron is negative, the opposite is the case, and the turn of the plane of polarization is then in the same direction as that of the positive current in a solenoid having its lines of force coincident with that of the field.

The right-hand side of (29) being inversely proportional to $(n^2 - \omega^2)^2$ becomes abnormally great when the period of the transmitted light approaches the free period of the molecule, but the direction of rotation remains the same whether ω is greater or smaller than n.

When n is nearly equal to ω, equation (29) fails to be correct, and we must derive r from (28).

172. Connexion between the Zeeman effect and magneto-gyration. We might have derived the results of the last article directly from the general theory of refraction, taking account of the changes in the free periods due to the magnetic field. Equation (42) Chapter XI. gives us for the velocity of a wave in a responsive medium:

$$\frac{1}{v^2} = K + \frac{4\pi e^2 N}{\rho(n^2 - \omega^2)} \quad \text{.......................(30)}.$$

In the magnetic field, the free period $2\pi/n$ is altered, and is different for circular vibrations according as they are left-handed or right-handed. According to (19) of this Chapter we must therefore substitute $\rho n^2 \pm He\omega$ for ρn^2, the upper sign holding for the right-handed rotation. This introduced into (30) leads directly to (27).

The investigation of the last article has been derived from an important paper by W. Voigt, who first gave equations which are practically identical with, though in one respect more general than (28). Voigt adds a frictional term to the equations of motion, in order to include the phenomenon of absorption, but owing to the objections raised in Art. 153 against the introduction of this term it has been omitted here.

The importance of Voigt's work consists in the establishment of a simple and rational connexion between the Zeeman effect and magneto-gyric properties. Each free period of the molecule is divided by the magnetic field into two, one being dextro-gyric, and the other laevo-gyric. Each of these imposes a rotatory polarization in its own direction, the velocity of propagation being increased on the violet side and diminished on the red side. Consider a period on the red side of a Zeeman doublet. It is most affected by the least refrangible component, the effect being a *diminution* of velocity, hence the resulting photo-gyric effect is in the same direction as that of the *most* refrangible component. On the violet side the most refrangible component is the one that is most active, and as the effect is here an *increase* in velocity, it follows that the photo-gyric effect is also in this case in the direction of the most refrangible component of the Zeeman doublet. This is true for all vibrations which do not fall within the periods intermediate between those of the two components, where the effect is in the opposite sense as easily reasoned out in the same manner. Zeeman's observations on Sodium light show that the most refrangible component rotates in the direction of the solenoidal current, giving a magnetic force coincident with that of the field, and this is therefore the direction in which we should expect sodium vapour to rotate the plane of polarization, except within a very narrow range close to the undisturbed period. Observation confirms this.

We conclude our theoretical discussion by deducing a remarkable relation first brought forward on more speculative grounds by H. Becquerel. Not necessarily confining ourselves to single free periods, we may write equation (30)

$$\frac{1}{v^2} = K + \Sigma \frac{\beta}{(n^2 - \omega^2)} \quad \text{.....................} (30a),$$

where the summation extends to the different values of n for which β may also have different values.

With the ordinary notation for small quantities, we may put

$$\delta \frac{1}{v^2} = \frac{d}{dn}\left(\frac{1}{v^2}\right)\delta n,$$

or

$$\frac{2\delta v}{v^3} = -\frac{d}{dn^2}\left(\frac{1}{v^2}\right)\delta n^2.$$

If $1/v^2$ has the form of (30a) we may substitute differentiation with respect to ω^2 for differentiation with respect to $-n^2$ and hence

$$\frac{\delta v}{v^2} = \tfrac{1}{2} v \frac{d}{d\omega^2}\left(\frac{1}{v^2}\right)\delta n^2.$$

If δv represents the increase in the velocity of propagation of the laevo-gyric light due to the magnetic field, the gyric coefficient (r) is $\omega\delta v/v^2$. For δn^2 we may write, according to (19) and (20), $4\pi H\omega z$, so that

$$r = \pi z H \omega v \frac{d}{d\omega}\left(\frac{1}{v^2}\right)$$

$$= 2\pi \frac{zH\omega}{V} \frac{d\mu}{d\omega}$$

$$= -2\pi \frac{zH\lambda}{V} \frac{d\mu}{d\lambda} \quad \text{..........................} (31),$$

where the refractive index μ has been substituted for V/v. This is Becquerel's equation*, which will be further discussed in the next article.

173. Experimental Facts and their connexion with the theory. The magneto-gyric effects of the great majority of substances are in the positive direction, by which we mean that they are in the same direction as that of the solenoidal current producing the magnetic field. If our theory is correct, this would mean that it is the negative electron which is the active vibrator, a result which we had already derived from the Zeeman effect. The salts of iron form however a

* *C. R.* cxxv. p. 679 (1897).

notable exception, for it is found that those salts which are magnetic have a negative coefficient. This at one time led to the belief that there might be a characteristic difference between dia-magnetic and para-magnetic bodies, the latter possessing a negative coefficient. The following table which has been given by H. du Bois shows how far such a distinction is justified.

TABLE XIV.

Diamagnetic		Paramagnetic	
Dextro-gyric	Laevo-gyric	Dextro-gyric	Laevo-gyric
Potassium ferro-cyanide	Titanium chloride	Iron	Ferrous salts
Lead borate		Cobalt	Ferric salts
Water		Nickel	Potassium ferricyanide
Hydrogen		Oxygen	Chromium trioxide
The great majority of solid, liquid and gaseous substances		Nitric oxide	Potassium bichromate
		Cobalt salts	Potassium chromate
		Nickel salts	Cerium salts
		Manganese salts	Lanthanum salts
		Cupric salts	Didymium salts

It is notable that the three magnetic metals, iron, nickel and cobalt, have a positive gyric coefficient, which seems at first sight in direct contradiction to the suggested connexion. But it has been found that for these metals $d\mu/d\lambda$ is positive, so that if Becquerel's law is generally true, the negative value of r might be explained. Titanium chloride is the only diamagnetic body which gives a negative r, but Titanium is a magnetic metal and therefore it is possible to argue that the dia-magnetism of chlorine overpowers the magnetism of Titanium, but that with regard to the gyric property the metal has the upper hand. The same argument cannot however be used to explain the positive coefficient of oxygen, and the salts of cobalt, nickel and manganese. The subject is suggestive, but requires further experimental treatment. Should the negative coefficients be ultimately found to be confined to magnetic substances, it will not be necessary to assume that their vibrating electrons are positive. The magnetic molecule may have a gyric property in virtue of its being magnetic, and the effects of this property would superpose themselves on the other effects, to which our theory has been confined. In the case of feebly magnetic substances, the Zeeman gyration may gain the upper hand, while in chloride of Titanium, the pure magnetic vortex rotation may be superposed to be paramount. This might explain some of the discrepancies of the above Table. A theory of magnetic vortex-gyration has been given by Drude*.

The following table given by H. Becquerel† shows the magnitude of the rotation of different substances compared with carbon bisulphide, and gives also the values of the Zeeman coefficient calculated from (31).

TABLE XV.

Substance	Relative magneto-gyric coefficient	$\lambda \frac{d\mu}{d\lambda}$	$z \times 10^{-5}$
Oxygen	·000146	$1·47 \times 10^{-5}$	5·98
Air	·000159	$1·44 \times 10^{-5}$	6·64
Nitrogen	·000161	$1·68 \times 10^{-5}$	5·74
Carbonic acid	·000302	$2·00 \times 10^{-5}$	9·07
Nitrous oxide	·000393	$4·85 \times 10^{-5}$	4·88
Water	·308	$1·99 \times 10^{-2}$	9·33
Benzine	·636	$4·88 \times 10^{-2}$	7·85
Phosphor trichloride	·651	$4·71 \times 10^{-2}$	8·30
Carbon bisulphide	1·000	$9·71 \times 10^{-2}$	6·20
Liquid phosphorus	3·120	$2·52 \times 10^{-1}$	7·41
Titanium bichloride	−1·358	$9·96 \times 10^{-2}$	−2·16

Excluding the dextro-gyric titanium bichloride it will be noticed that the Zeeman coefficients for these substances, having widely different dispersions, are all of the same order of magnitude, thus giving a substantial confirmation of the correctness of Becquerel's law. The average value of z in the above Table is about half that of the lowest and one quarter that of the highest number obtained directly from observations with luminous vapours. An apparent increase of electric mass, in the more complicated structures of molecules which do not give line spectra, is thereby suggested.

Absolute determinations of the magnetogyric coefficients have been made for carbon bisulphide and for water. The most recent determinations, reduced to unit magnetic force, are, in minutes of arc :

* *Lehrbuch der Optik*, p. 384.
† *C. R.* cxxv. p. 683 (1897).

TABLE XVI.

Bisulphide of Carbon.

Lord Rayleigh*	for Sodium light,		$t=18°$,	$r=0'{\cdot}04200$
Kopsel†	,,	,,	,,	$0'{\cdot}04199$
,,	,,	,,	$t=0°$,	$0'{\cdot}04207$
Becquerel‡	,,	,,	,,	$0'{\cdot}04341$

Water.

Arons§	$t=23°$,	$0'{\cdot}01295$

The gyric effect of thin films of iron, when magnetized to saturation, is enormous. Its discoverer, Kundt, found it to be at the rate of a complete revolution for a thickness of ·02 mm. which gives 200,000° for one centimetre, an effect which is 290 million times greater than in bisulphide of carbon. Cobalt gives a value nearly as great as found by Du Bois, and nickel about half as great.

The magneto-gyric coefficient is in general roughly proportional to the square of the frequency, showing that the vibrations which chiefly determine it have a very small wave-length, but anomalous cases have been noted. Kundt‖ observed that thin iron films rotate the plane of polarization of red light more than that of blue light, and Lobach¶ measured the rotational coefficients of iron, nickel and cobalt in different parts of the spectrum. The diminution in the angle of rotation between $\lambda = 6{\cdot}7 \times 10^{-5}$ and $\lambda = 4{\cdot}3 \times 10^{-5}$ was found to be approximately for iron 45 %, for cobalt 23 %, and for nickel 41 %.

An interesting confirmation of the theory given in Art. 169 is obtained by the observation of the gyric effects in the neighbourhood of absorbing regions of the spectrum. As has been pointed out in that article, the introduction of the Zeeman effect into Sellmeyer's equation leads directly to the conclusion that on both sides of an absorption line, there is a strong magneto-gyric effect in the direction in which the more refrangible members of the Zeeman components rotate. This fact had been observed in the neighbourhood of the sodium lines by Macaluso and Corbino** (the gyric coefficient being positive, *i.e.* the rotation in the direction of the current producing the field). It has been further extended and commented upon by H. Becquerel††.

* *Collected Works*, Vol. II. p. 360.

† *Wied. Ann.* Vol. XXVI. p. 456, 1885.

‡ *Ann. Chim. Phys.* Vol. XXVII. p. 312, 1882.

§ *Wied. Ann.* Vol. XXIV. p. 161, 1885.

‖ *Wied. Ann.* Vol. XXIII. p. 237, 1884.

¶ *Wied. Ann.* Vol. XXXIX. p. 346, 1890.

** *C.R.* CXXVII. p. 548, 1898.

†† *C.R.* CXXVII. p. 647, 1898.

174. Double Refraction at right angles to the lines of force. If a plane wave traverses the magnetic field at right angles to the lines of force, the vibrations parallel to the field are propagated with different velocities from those at right angles. This follows also from Sellmeyer's theory of refraction in combination with the Zeeman effect.

Taking the simplest case of a line split by the magnetic field into a Zeeman triplet, the outer components affect the velocity of light of the vibrations normal to the field, while the central component affects the vibrations parallel to the field. Approaching an absorption line from the less refrangible side, the first effect will be a diminution of the velocity of both components, but to a greater degree of that component which lies nearest, *i.e.* the vibration normal to the field. Similarly approaching the absorption line from the violet end, both components are accelerated, and it is again the component vibrating normally to the field which is most affected. Hence there is double refraction in such a sense that towards the red end the vibration parallel to the line of force is propagated most quickly and on the violet side the vibration normal to the field. This result was predicted by W. Voigt from the theory and verified experimentally by him in conjunction with Wiechert*.

To obtain an expression for the amount of double refraction to be expected, we write Sellmeyer's equation for the light vibrating normally to the field in the form:

$$\frac{1}{v_n^2} = K + \frac{\beta'}{(n_r^2 - \omega^2)} + \frac{\beta'}{(n_l^2 - \omega^2)},$$

where $$n_r^2 = n^2 + 4\pi z\omega H; \quad n_l^2 = n^2 - 4\pi z\omega H;$$

$$\therefore \frac{1}{v_n^2} = K + \frac{2\beta'(n^2 - \omega^2)}{(n^2 - \omega^2)^2 - 16\pi^2 z^2\omega^2 H}.$$

The vibrations parallel to the lines of force are undisturbed and hence

$$\frac{1}{v_p^2} = K + \frac{\beta}{n^2 - \omega^2}.$$

For $H = 0$, the two expressions must agree, and hence $2\beta' = \beta$. Writing α^2 for

$$16\pi^2 z^2\omega^2 H^2/(n^2 - \omega^2),$$

we have

$$\frac{1}{v_n^2} - \frac{1}{v_p^2} = \frac{\beta}{n^2 - \omega^2 - \alpha^2} - \frac{\beta}{n^2 - \omega^2}$$

$$= \frac{\beta\alpha^2}{(n^2 - \omega^2)(n^2 - \omega^2 - \alpha^2)}.$$

* *Wied. Ann.* Vol. LXVII. p. 345, 1899.

If we treat α^2 as a small quantity compared with $n^2 - \omega^2$ and reintroduce its value, we find

$$\frac{1}{v_n^2} - \frac{1}{v_p^2} = \frac{16\beta\pi^2 z^2 \omega^2 H^2}{(n^2 - \omega^2)^3}.$$

When the vibrations on the contrary are so near to the free undisturbed period that $n^2 - \omega^2$ is small compared with α^2:

$$\frac{1}{v_n^2} - \frac{1}{v_p^2} = -\frac{\beta}{(n^2 - \omega^2)}.$$

The double refraction is now in the opposite direction, and hence close to the free period the perpendicular vibrations are propagated more quickly when $n > \omega$, *i.e.* on the side of lower frequency.

CHAPTER XIII.

TRANSMISSION OF ENERGY.

175. Propagation of Energy. Energy may be transmitted through a surface either by the passage of matter in motion, or by the performance of work. An example of the first kind of transference of energy is furnished by the conduction of heat through gases, the kinetic energy carried by the molecules through a surface at right angles to the flow of heat, being greater in the direction of the flow than in the reverse direction. But we are not here concerned with this simple and direct method of transference of energy.

Waves propagated through elastic solids carry energy across a surface owing to the tangential forces, which will in general do work. A transference of energy results though the velocities and stresses are alternately in opposite directions, when their product contains a part which is not periodic. As the propagation of energy can only be accurately investigated when the mechanism of the motion is known we study in the first instance some simple cases of the transmission of waves through elastic bodies.

176. Waves of pure compression or dilatation in a perfect fluid. We choose a frictionless fluid in order to simplify the equations as much as possible. Putting $n=0$, and writing ξ, η, ζ for the displacements, equations (9) Art. 132 become

$$\left.\begin{aligned}\frac{d^2\xi}{dt^2} &= v^2\frac{d\delta}{dx}\\ \frac{d^2\eta}{dt^2} &= v^2\frac{d\delta}{dy}\\ \frac{d^2\zeta}{dt^2} &= v^2\frac{d\delta}{dz}\end{aligned}\right\}\text{.............................(1)},$$

here v^2 is written for κ/ρ and

$$\delta = \frac{d\xi}{dx} + \frac{d\eta}{dy} + \frac{d\zeta}{dz} \quad \text{......................(2).}$$

For the stresses we have according to (7) and (8) Art. 131

$$S = T = U = 0,$$
$$P = Q = R = v^2 D\delta,$$

where the density is now represented by D.

If we consider in the first place a plane wave, the displacements being parallel to the axis of x, the first of equations (1) becomes

$$\frac{d^2\xi}{dt^2} = v^2 \frac{d^2\xi}{dx^2},$$

which is satisfied by

$$\xi = A \sin \frac{2\pi}{\lambda}(x - vt).$$

If W represents the work done across the surface

$$\frac{dW}{dt} = -P\frac{d\xi}{dt}$$
$$= -v^2 D \frac{d\xi}{dx} \cdot \frac{d\xi}{dt},$$

where the negative sign has to be introduced on the right hand because P is taken as positive when it is a *tension* and acts therefore in the opposite direction to that in which the velocities are taken as positive.

Substituting for ξ and confining our attention to the plane $x = 0$,

$$\frac{dW}{dt} = \frac{4\pi^2 v^3 DA^2}{\lambda^2} \cos \frac{2\pi}{\lambda} vt.$$

By integration

$$W = \frac{2\pi^2 v^3 DA^2}{\lambda^2}\left(t + \frac{\lambda}{4\pi v} \sin \frac{4\pi}{\lambda} vt\right).$$

The second term vanishes at intervals of time which are equal to half a complete period, and becomes more and more negligible as t increases. Leaving this term out of account, we may write for the work transmitted through unit surface

$$W = \tfrac{1}{2} D V_1^2 vt \quad \text{..............................(3),}$$

where V_1 stands for the maximum velocity. If the whole mass of air through which the waves have spread in time t had a velocity equal to the maximum velocity V_1, its kinetic energy would be equal to that transmitted through the surface. As the average kinetic energy in a simply periodic wave is equal to half the maximum energy, only half the energy transmitted through the surface is in the kinetic form, the other half being potential.

It is important to notice that the transmission of energy depends on the coincidence of the phases of velocity and pressure. The condensed portions of the fluid move in the direction in which the wave is propagated, and the rarefied portions in the opposite direction ; hence the work done while the air moves forwards is not undone while the air moves backwards.

We next take the case of waves diverging from a point. The motion to be considered belongs to an important class in which the velocities may be represented as the partial differential coefficients of the same function ϕ, called the velocity potential.

Put $$\frac{d\xi}{dt}=\frac{d\phi}{dx};\quad \frac{d\eta}{dt}=\frac{d\phi}{dy};\quad \frac{d\zeta}{dt}=\frac{d\phi}{dz}.$$

Equations (1) are now all contained in the simple equation

$$\frac{d^2\phi}{dt^2}=v^2\nabla^2\phi \quad \text{..........................(4).}$$

For the stress P we have

$$\begin{aligned}\frac{dP}{dt}&=v^2D\frac{d\delta}{dt}\\ &=v^2D\nabla^2\phi\\ &=D\frac{d^2\phi}{dt^2}.\end{aligned}$$

Hence if P be zero in a state of rest:

$$P=D\frac{d\phi}{dt}.$$

If ϕ depends only on the distance r from a fixed point which acts as a source from which the vibrations emanate, we have

$$\frac{d\phi}{dx}=\frac{d\phi}{dr}\frac{x}{r},$$

$$\frac{d^2\phi}{dx^2}=\frac{d^2\phi}{dr^2}\frac{x^2}{r^2}-\frac{d\phi}{dr}\cdot\frac{1}{r}-\frac{d\phi}{dr}\cdot\frac{x^2}{r^3}.$$

Changing similarly the variable in $\frac{d^2\phi}{dy^2}$, $\frac{d^2\phi}{dz^2}$, we find by addition

$$\begin{aligned}\nabla^2\phi&=\frac{d^2\phi}{dr^2}+\frac{2}{r}\frac{d\phi}{dr}\\ &=\frac{1}{r}\frac{d^2r\phi}{dr^2}.\end{aligned}$$

Equation (4) now becomes

$$\frac{d^2r\phi}{dt^2}=v^2\frac{d^2r\phi}{dr^2},$$

the solution of which is $\qquad r\phi = f(r - vt)$,
or confining ourselves to the simple periodic motion

$$\phi = \frac{A}{r} \sin \frac{2\pi}{\lambda} (r - vt).$$

This value of ϕ is therefore a solution of the differential equation (4) For the radial velocity we have by differentiation

$$\frac{d\phi}{dr} = \frac{A}{r} \left\{ \frac{2\pi}{\lambda} \cos \frac{2\pi}{\lambda} (r - vt) - \frac{1}{r} \sin \frac{2\pi}{\lambda} (r - vt) \right\} \dots\dots\dots (5).$$

The volume of fluid passing in unit time through the concentric spherical surfaces round the origin as centre is $4\pi r^2 d\phi/dr$, which near the origin where r is small compared with λ becomes equal to $4\pi A \sin \frac{2\pi}{\lambda} vt$. Our equations therefore represent a state in which an elastic fluid is alternately introduced and withdrawn at the origin, the consequent changes of pressure being then transmitted outwards in the form of condensational waves. It is instructive to compare the two terms the sum of which according to (5) makes up the velocity. The first, varying inversely as the distance, is the important one at distances great compared with the wave-length, the second varying inversely as the square of the distance is only appreciable near the origin. We note the difference in phase between the two terms amounting to a right angle, with which we have already become familiar in the elementary treatment of wave-propagation (Arts. 47 and 48). Equation (5) further shows that in consequence of the second term the energy of motion near the origin diminishes very rapidly. As regards the energy transmitted outwards we know that in a progressive wave and in the absence of any dissipation through absorption or otherwise, the law of the inverse square must hold. We conclude that near the origin the energy must partly be stationary. We can prove that the transmitted energy is in accordance with what is to be expected from the laws of energy, by remembering that the work done per unit time across unit surface is equal to the product of the pressure and the component of the velocity which is normal to the surface. In the case considered, the velocity being everywhere radial, the rate of work transmitted through a sphere of radius r is

$$\begin{aligned} dW/dt &= -4\pi r^2 P\, d\phi/dr \\ &= -4\pi D r^2\, d\phi/dt \,.\, d\phi/dr \\ &= 2\pi D A^2 \sin(\omega t - lr) \left\{ l \sin(\omega t - lr) - \frac{1}{r} \cos(\omega t - lr) \right\}. \end{aligned}$$

Integrating with respect to the time and leaving out periodic terms, we find $\qquad W = 2\pi\omega D A^2 l t.$

This expression does not contain r and hence the work transmitted through concentric spheres enclosing the origin is constant. It follows that the work transmitted in a given time through unit surface varies inversely as the square of the distance. For a fuller treatment of the subject the reader is referred to Lord Rayleigh's treatise on Sound, Vol. II., Arts. 279 and 280.

177. Plane waves of distortion in an elastic medium. Let the displacements be parallel to the axis of z and be denoted by ζ, the wave normal being the axis of x. The only force which can do work across the plane xy is the tangential stress which in Art. 129 has been called T, and which according to (7) Art. 131 is equal to $nd\zeta/dx$, ξ being zero in the present case. The stress is here taken to be positive when the portion of matter on the positive side of the plane yz acts on the matter which is on the negative side with a force directed along the positive axis of z. Hence for waves travelling in the positive direction, if W be the energy transmitted across unit surface,

$$\frac{dW}{dt} = -n\frac{d\zeta}{dx}\frac{d\zeta}{dt}.$$

If
$$\zeta = A \sin\frac{2\pi}{\lambda}(x - vt)$$

and the coefficient of distortion n is replaced by v^2/D, we find, as in the case of the sound-wave, leaving out periodic terms,

$$W = \tfrac{1}{2}DV_1^2 vt \quad \text{.........................}(6),$$

where V_1 denotes the maximum velocity.

178. Sphere performing torsional oscillations in an elastic medium. Consider displacements in an elastic medium defined by

$$\xi = 0, \quad \eta = -\frac{d\phi}{dz}, \quad \zeta = \frac{d\phi}{dy}$$

where

$$\phi = \frac{A}{r}\sin\frac{2\pi}{\lambda}(r - vt).$$

The displacements satisfy the condition

$$\frac{d\xi}{dx} + \frac{d\eta}{dy} + \frac{d\zeta}{dz} = 0,$$

which shows that there is no condensation.

Also
$$\begin{aligned}\nabla^2\eta &= -\frac{d}{dz}\nabla^2\phi \\ &= -\frac{1}{v^2}\frac{d}{dz}\cdot\frac{d^2\phi}{dt^2} \\ &= \frac{1}{v^2}\frac{d^2\eta}{dt^2}.\end{aligned}$$

Similarly $$\nabla^2\zeta = \frac{1}{v^2}\frac{d^2\zeta}{dt^2}.$$

It has been shown in Art. 135 that these equations satisfy all the conditions of a wave propagation, the waves being purely distortional, and v^2 being identified with n/D.

As ϕ does not contain x, y, z explicitly,

$$\frac{d\phi}{dy} = \frac{d\phi}{dr} \cdot \frac{y}{r}, \quad \frac{d\phi}{dz} = \frac{d\phi}{dr} \cdot \frac{z}{r}.$$

Hence $$x\xi + y\eta + z\zeta = -y\frac{d\phi}{dz} + z\frac{d\phi}{dy}$$
$$= 0.$$

It follows that the displacements at any point are at right angles to the radius vector drawn from the origin to that point. As there are no displacements parallel to the axis of x, the displacements are along circles drawn round OX as axis.

Let ρ be the distance of any point from the axis, so that $r^2 = x^2 + \rho^2$. We obtain the amount of the displacement by resolving η and ζ in a direction at right angles to ρ in a plane parallel to the plane of yz. This gives for the displacement:

$$\frac{-\eta z + \zeta y}{\rho} = \frac{1}{\rho}\left(z\frac{d\phi}{dz} + y\frac{d\phi}{dy}\right)$$
$$= \frac{z^2 + y^2}{\rho r}\frac{d\phi}{dr}$$
$$= \frac{\rho}{r}\frac{d\phi}{dr}.$$

The angular displacement obtained by dividing the actual displacement by ρ only depends on r, and is therefore the same at all points of a sphere having the origin as centre. Each such sphere performs torsional oscillations as if it were rigid. We may therefore imagine an inner sphere to be actually rigid and the oscillations to be maintained by forces applied to this sphere. Our system of equations will then tell us how these oscillations are propagated outwards.

In the language of Optics the vibrations at any point are polarized in a plane passing through OX which is the axis of rotation. The angular displacements are

$$\frac{1}{r}\frac{d\phi}{dr} = \frac{2\pi A}{\lambda r^2}\cos\frac{2\pi}{\lambda}(r - vt) - \frac{A}{r^3}\sin\frac{2\pi}{\lambda}(r - vt) \quad \text{.........(7)},$$

and are nearly equal to the first or second term of this expression respectively, according as r is very small or very large compared with $\lambda/2\pi$. Comparing large and small values of r, we have here the same change of phase of a right angle which has been noted in Art. 173.

The maximum angular displacement at a distance S from the origin as obtained from (7) is:

$$A\sqrt{4\pi^2S^2+\lambda^2}/\lambda S^3,$$

and this is the amplitude of oscillation which must be maintained at a sphere of radius S in order to cause an angular amplitude $2\pi A/\lambda r^2$ at a large distance. If the maintained angular amplitude is B, it follows that for large distances the angular amplitude is

$$\frac{BS^2}{r^2\sqrt{1+\lambda^2(4\pi^2S^2)^{-1}}}.$$

The actual amplitude is obtained on multiplying this expression by $r\sin\theta$, where θ denotes the angle which r forms with OX. To calculate the energy communicated by the rigid sphere to the surrounding medium, we make use of the obvious proposition that the energy transmitted through all concentric spheres must be equal and we may therefore simplify the calculation by considering only a sphere of very large radius.

If we write $(V_1\sin\theta)/r$ for the maximum velocity at a large distance, the total energy transmitted through unit surface at any time is by (6)

$$W=\tfrac{1}{2}DV_1^2vt\sin^2\theta/r^2,$$

and the work transmitted through the complete sphere is

$$\int_0^\pi 2\pi Wr^2\sin\theta\,d\theta=\int_0^\pi \pi tDV_1^2v\sin^3\theta\,d\theta=\frac{4\pi}{3}DV_1^2vt \quad\text{......(8).}$$

Substituting the value of V_1, we find for E, the total energy transmitted,

$$E=\frac{4\pi Dvt}{3\lambda}\,\frac{S^4}{1+\lambda^2(4\pi^2S^2)^{-1}}\left(\frac{2\pi Bv}{\lambda}\right)^2$$

$$=\frac{4\pi}{3}\,\frac{DSvt}{1+\lambda^2(4\pi^2S^2)^{-1}}\left(\frac{2\pi BSv}{\lambda}\right)^2.$$

The bracket on the right-hand side represents the greatest velocity in the equatorial plane of the rigid sphere.

It should be noticed that the energy transmitted diminishes with increasing wave-length (*i.e.* increasing period) and this diminution is the more important the smaller the radius of the embedded sphere is compared with the wave-length.

179. Waves diverging from a sphere oscillating in an elastic medium. The problems discussed in this and the preceding article were first solved by W. Voigt*. Kirchhoff† considerably simplified the mathematical analysis and more recently Lord Kelvin‡

* *Crelle's Journal*, Vol. LXXXIX. p. 288.

† *Crelle's Journal*, Vol. XC. p. 34.

‡ *Phil. Mag.* Vol. XLVII. p. 480 and XLVIII. pp. 277, 388.

has treated the same question very completely, adding several new and interesting results. We imagine a sphere embedded in an elastic medium, to which it is rigidly attached, and performing periodic linear oscillations according to the formula $A \sin \omega t$. The reader is referred to Lord Kelvin's *Baltimore Lectures* for the complete solution of the problem, we shall here confine ourselves to the question of emission of energy in an incompressible medium. For this purpose it is only necessary to consider the motion at a distance which is large compared with the radius of the sphere (S) and the wave-length. If the oscillations of the sphere take place along the axis of x, Kelvin's equations for the displacement when r is very great, are

$$\left.\begin{aligned} \xi &= -\frac{3}{2} AS\left(\frac{x^2}{r^3} - \frac{1}{r}\right)\sin(\omega t - lr) \\ \eta &= -\frac{3}{2} AS\frac{xy}{r^3}\sin(\omega t - lr) \\ \zeta &= -\frac{3}{2} AS\frac{xz}{r^3}\sin(\omega t - lr) \end{aligned}\right\} \text{..............(9).}$$

These equations give:

$$\xi x + \eta y + \zeta z = 0,$$

showing that the vibrations take place at right angles to the radius vector. The symmetry of the expression for the displacements as regards y and z shows that the displacements take place in meridional planes For the resultant oscillation we have

$$(\xi^2 + \eta^2 + \zeta^2)^{\frac{1}{2}} = \frac{3}{2r} AS \sin\theta \sin(\omega t - lr),$$

where θ denotes the angle between the radius vector and the axis of x. The sign of the square root which occurs on the left side is determined by the consideration that when $x = 0$, the last equation must agree with the first of the equations (9). We note that in this case there is not the change of phase of a right angle which occurs when the sphere performs torsional oscillations. In the language of Optics, the sphere may be said to send out polarized light, the vibrations being in meridional planes and at right angles to the ray. The amplitude is a maximum in the equatorial plane, zero along the axis, and in intermediate positions is proportional to $\sin\theta$. If as in the last article, we write $V_1 \sin\theta/r$ for the maximum velocity, we may apply (8) directly to obtain the transmitted energy which is

$$\begin{aligned} E &= \frac{4\pi}{3} D V_1^2 vt \\ &= 3\pi D\omega^2 A^2 S^2 vt. \end{aligned}$$

The emission of energy is therefore inversely proportional to the square of the period.

180. Divergent Waves of Sound. The theory of Sound furnishes several important applications of the communication of energy from a vibrating body to a surrounding medium. If a stretched string vibrates backwards and forwards, the air which is compressed on one side is able to flow round the string and to diminish the rarefaction which tends to form behind the string. Under these circumstances comparatively little energy escapes in the form of sound-waves. Stokes* calculated the emission of the actual sound and compared it with that which would have been emitted if the lateral motion in the neighbourhood of the string were omitted. For a piano string of ·02 inch radius sounding the middle C (wave-length about 25 inches) it appears that the prevention of the lateral motion would increase the intensity 40,000 times. This, as Stokes points out, shows the importance of sounding-boards, the broad surface of which is able to excite intense vibrations even though the motion itself is small. The following experiment may be described in Stokes' own words. "The increase of sound produced by the stoppage of lateral motion may be prettily exhibited by a very simple experiment. Take a tuning-fork, and holding it in the fingers after it has been made to vibrate, place a sheet of paper, or the blade of a broad knife, with its edge parallel to the axis of the fork, and as near to the fork as conveniently may be without touching. If the plane of the obstacle coincide with either of the planes of symmetry of the fork, as represented in section at A or B, no effect is produced; but if it be placed in an intermediate position, such as C, the sound becomes much stronger."

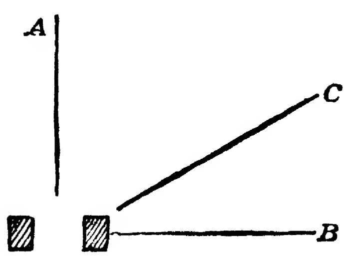


Fig. 177.

The motion of air round the sounding body is the more effective the shorter the wave-length. Were the length of the wave infinitely great, the air would move like an incompressible fluid backwards and forwards round the source of sound, and there would be no emission of energy once this motion is established. Stokes shows by applying the analysis to the case of vibrating spheres, that this is the explanation of an experiment due to Leslie, in which the sound of a bell placed in a partially exhausted receiver is diminished by the introduction of hydrogen.

181. Transmission of energy by electromagnetic waves. When we consider the transmission of energy in a variable electromagnetic field we are met at once by the difficulty that we are ignorant

* *Phil. Trans.* Vol. CLVIII. p. 447 (1868), Rayleigh, *Theory of Sound*, Vol. II. p. 306.

of the mechanism by which electromagnetic action is propagated. Hence we cannot obtain an expression for the work done across a surface, unless we form an hypothesis to specify the displacements in the medium. In the special cases which we have to consider here, the difficulty may be turned. According to Maxwell a medium of permeability μ, subject to a magnetic force which at any place is H, possesses energy which per unit volume is measured by $\mu H^2/8\pi$. Similarly if K be the dielectric constant and E the electric force, $KE^2/8\pi$ is the electric energy also per unit volume. Consider now a plane wave propagated with velocity v in the direction of the axis of z.

Equations (22) Art. 140, which determine P, Q, the components of electric force, in terms of α, β, the components of magnetic force, give us

$$KE^2 = K(P^2 + Q^2) = K\mu^2 v^2 (\alpha^2 + \beta^2).$$

As $K\mu = 1/v^2$ and $\alpha^2 + \beta^2$ measures the square of magnetic force, it follows that

$$KE^2 = \mu H^2;$$

so that we may, in the case considered, write for the energy per unit volume either $\mu H^2/4\pi$ or $KE^2/4\pi$. The wave need not be homogeneous and may be either plane or elliptically polarized or not polarized at all. Consider now a wave-front advancing from left to right and coinciding with the plane of xy at the time $t = 0$. At any time t, the wave-front will be at a distance vt from the origin, and the energy which has crossed unit surface of the plane of xy will be that contained in the volume having the unit surface as base and as length the distance vt measured along the axis of z. If the magnetic force in the wave-front is of the form

$$H = H_0 \cos(\omega t - lz),$$

the average value of H^2 is equal to $\frac{1}{2}H_0^2$. Hence the energy which has crossed unit surface in time t, putting $\mu = 1$, is $H_0^2 vt/8\pi$. The work done across a small surface of any wave-front cannot depend on the question whether the wave is plane or not. We are therefore justified in using the expression obtained whenever the electromagnetic disturbance follows the simple periodic law.

We next treat of a simple case in which we can trace the loss of energy of a radiating source. We adapt for this purpose the results of Art. 178, substituting the magnetic force for the displacement, so that we may write

$$\alpha = 0, \quad \beta = -\frac{d\phi}{dz}, \quad \gamma = \frac{d\phi}{dy} \quad \ldots\ldots\ldots\ldots\ldots\ldots(10),$$

where

$$\phi = \frac{A}{r}\sin(lr - \omega t) \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(11).$$

These equations satisfy (19) of Art. 138, and represent therefore a possible distribution of magnetic force. It follows from the results of Art. 178 that the magnetic force at a distance ρ from the axis of x is $\rho \frac{d\phi}{dr} \Big/ r$ and that the lines of magnetic force are circles having OX as axis.

We next consider a region round the origin so small that the phase at all points lying in it may be considered identical with that at the origin. Within this region we may write

$$\phi = -\frac{A}{r} \sin \omega t.$$

The components of current are obtained from (12) Art. 136:

$$\begin{aligned} 4\pi u &= \frac{d\gamma}{dy} - \frac{d\beta}{dz} = \frac{d^2\phi}{dy^2} + \frac{d^2\phi}{dz^2} \\ &= -\frac{d^2\phi}{dx^2}, \\ 4\pi v &= -\frac{d^2\phi}{dydx}, \\ 4\pi w &= -\frac{d^2\phi}{dzdx}. \end{aligned}$$

For the electric forces as obtained from (17), Art. 137, we have

$$\begin{aligned} KP &= -\frac{d}{dx} \cdot \frac{d^{-1}}{dt} \cdot \frac{d\phi}{dx} = \frac{d}{dx} \cdot \frac{Ax}{\omega r^3} \cos \omega t, \\ KQ &= -\frac{d}{dy} \cdot \frac{d^{-1}}{dt} \cdot \frac{d\phi}{dx} = \frac{d}{dy} \cdot \frac{Ax}{\omega r^3} \cos \omega t, \\ KR &= -\frac{d}{dz} \cdot \frac{d^{-1}}{dt} \cdot \frac{d\phi}{dx} = \frac{d}{dz} \cdot \frac{Ax}{\omega r^3} \cos \omega t. \end{aligned}$$

The electric forces close to the origin are therefore derivable from a potential

$$-\frac{Ax}{\omega r^3} \cos \omega t.$$

If a quantity of electricity $-e$ is placed on the axis of x at a distance $\frac{1}{2}h$ from the origin, and similarly a quantity $+e$ at the same distance on the negative side, the electrostatic potential of such a so-called doublet is known to be $-ehx/r^3$. Hence we may represent the electric forces in the case we are considering by means of a doublet if we make

$$eh = (A \cos \omega t)/\omega.$$

The system of waves represented by (10) and (11) may therefore be considered to be produced by a vibrating electric doublet at the origin, the two charges oscillating in the same period, reaching a maximum distance h and crossing at the origin. The quantity eh is called the moment of the doublet, for which we may write $M \cos \omega t$.

The total energy dissipated per unit time by the vibrations of such a doublet may now be calculated. At a large distance the magnetic forces are

$$\beta = \frac{lz}{r^2} M\omega \cos(lr - \omega t),$$

$$\gamma = -\frac{ly}{r^2} M\omega \cos(lr - \omega t),$$

from which we obtain;

$$\beta^2 + \gamma^2 = l^2\omega^2 M^2 \sin^2\theta \cos^2(lr - \omega t)/r^2 \quad \text{...........(12)},$$

where θ is the angle between the radius vector and the axis of x. Hence through unit surface of large spheres the amount of energy which passes in time t is $l^2\omega^2 M^2 \sin^2\theta\, vt/8\pi r^2$. Integrating this over the whole sphere it becomes $\frac{1}{3}l^2\omega^2 M^2 vt$, or expressing l and ω in terms of τ and v (where τ represents the time of vibration), we finally obtain for the energy sent out in time t by the vibrating doublet:

$$16\pi^4 M^2 t/3v\tau^4 = M^2\omega^4 t/3v.$$

In order to form some numerical estimate of this loss of energy consider the positive charge to remain fixed at the origin, and the negative charge to vibrate according to the law $h \cos \omega t$. The maximum energy of the negative electron is $\frac{1}{2}m\omega^2 h^2$ or $\frac{1}{2}m\omega^2 M^2/e^2$, if m denotes its mass. From this we calculate the fraction of the maximum energy which is lost in a complete vibration taking up a time $2\pi/\omega$, to be $8\pi^2 e^2/3\lambda m$. The ratio e/m is known to be nearly $1{\cdot}8 \times 10^7$ and for e we may substitute 10^{-20}. This gives the loss of energy as being: $4{\cdot}8 \times 10^{-12}/\lambda$. For violet light we have $\lambda = 4 \times 10^{-5}$, so that in each period a particle sending out such light would lose about the ten-millionth part of its energy. The motion of the particle, taking account of the loss of energy by radiation, would have to be represented by the expression $he^{-\kappa t} \cos \omega t$, where the coefficient κ may be calculated from the data obtained. At each vibration there is a fractional diminution of the maximum velocity equal to $\kappa\tau$ and a fractional diminution of the energy equal to $2\kappa\tau$. Hence $1/\kappa\tau$ which is the number of vibrations in which the amplitude diminishes in the ratio $1 : e$ is $10^6/6$ or approximately 170,000. The small diminution in vibratory energy consequent on radiation justifies the criticism made in previous articles respecting the introduction of a frictional term in order to account for the so-called anomalous dispersion.

Consider now two similar doublets with their axes at right angles to each other, the positive electron being stationary and the negative oscillating in one case according to the law $h \cos \omega t$ and in the other according to the law $h \sin \omega t$. The electromagnetic effect will be the same as that of a single electron revolving with uniform speed in a circle of radius h, the loss of energy for each of the two vibrations at right angles to each other is that given above, and hence the total loss per unit time of such an electron revolving in time τ is $32\pi^4 M^2/3v\tau^4$. We arrive therefore at the remarkable conclusion that an electric charge describing a circle with uniform speed radiates energy, and as the speed is constant, the radiation can only depend on the acceleration which is directed to the centre. Writing f for the acceleration we have

$$f = \omega^2 h = \frac{4\pi^2 h}{\tau^2},$$

$$\therefore\ f^2 = \frac{16\pi^4 h^2}{\tau^4} = \frac{16\pi^4 M^2}{e^2\tau^4}.$$

We obtain therefore $\frac{2}{3}e^2 f^2/v$ for the loss of energy in unit time.

This expression which is here proved for the special case that the acceleration is at right angles to the motion holds generally so long as the velocity is small compared with the velocity of light. For a more detailed discussion, the reader is referred to Larmor's *Aether and Matter*, Chapter XIV.

182. Molecular scattering of light. If a source of light pass over a molecule or any particle which is small compared with the wave-length, a secondary disturbance spreads out from the small body. If the incident light be polarized, the scattered light is polarized also, the vibrations taking place in the plane passing through the direction of vibration at the obstacle; the amplitude of the scattered light is proportional to $\sin\theta$, if θ denote the angle between the vibration at the origin and the scattered ray. Although the intensity of the scattered light can only be calculated when we have defined the property of the obstacle in virtue of which it acts, some important conclusions may be drawn which are independent of any particular theory. It is well therefore to proceed as far as we can in the most general manner.

Let the original light be defined by some vector $R_0 \cos(\omega t - lx)$ where x is measured from the obstacle. The corresponding vector in the scattered light at a distance r may then be written in the form

$$[A_1 \cos(\omega t - lr) + B_1 \sin(\omega t - lr)]\, R_0 \sin\theta/r \quad\text{.........(13)},$$

where the factor B_1 indicates a possible change of phase at the obstacle. Taking the square of the amplitude as a measure of the energy and noting that the average value of $\sin^2\theta$ over the whole sphere is $2/3$, we find that the total energy emitted is $\frac{8}{3}\pi(A_1^2 + B_1^2) R_0^2$, and this must be derived from the energy of the incident beam. Let now a parallel pencil of light traverse in the direction of x a thin layer of a medium containing N particles per cubic centimetre. If E be the energy transmitted per unit surface we may define the coefficient k which measures the gradual weakening of the incident beam by $E^{-1}dE = kdx$. A layer of thickness dx scatters energy $\frac{8}{3}\pi (A_1^2 + B_1^2) R_0^2 N dx$ per unit surface, if R_0^2 is the energy per unit surface transmitted by the original beam. Hence

$$\tfrac{8}{3}\pi (A_1^2 + B_1^2) N = k \quad\text{.......................(14).}$$

We next find the effect of the layer dx at some point O which we take to be at a distance p from the layer. Drawing Fresnel's zones with O as centre and applying the method of Art. 46, we find that the total effect of the layer is equal in magnitude to half that of the first zone which has an area $\pi p\lambda$. The phase of the resulting vibration is that corresponding to a distance $p + \frac{1}{4}\lambda$ and the scattered light at O is therefore represented by

$$[A_1 \sin(\omega t - lp) - B_1 \cos(\omega t - lp)]\, R_0 \lambda N dx \quad\text{.........(15),}$$

the factor $2/\pi$ having been applied to account for the inequalities of phase of the vibration reaching O from different parts of the first zone (see Art. 46).

It should be noticed that it is only in the direction of the original light that the scattered rays can combine together so as to take part in the regular wave motion, because a displacement of one of the obstacles has no effect on the ultimate phase of the disturbance at a given point if the scattered ray lies in the direction in which the wave is propagated. The vibration indicated by (15) has now to be combined with that of the original wave, which at the point considered is $R_0 \cos(\omega t - lp)$. The term in B indicates a diminution in amplitude, while the term in A indicates a change of phase. If the layer dx were uniform as to its properties and had a refractive index μ, it would cause a retardation of the incident light $(\mu - 1)\, dx$ which is equivalent to an alteration in phase of $2\pi(\mu - 1)\, dx/\lambda$. We conclude that as regards the rate of propagation the medium acts like one having refractive index μ if

$$2\pi(\mu - 1) = \lambda^2 A_1 N \quad\text{.......................(16).}$$

The proportional diminution in amplitude is $B_1 \lambda N dx$ and the proportional diminution in energy will have twice that value. Hence with the same meaning of k as before

$$k = 2\lambda^2 B_1 N \quad\text{.............................(17).}$$

In all cases of more immediate interest B_1 is small compared with A_1, so that in (14) B_1^2 may be neglected. Combining that equation with (16) we obtain

$$k = \frac{32\pi^3 (\mu - 1)^2}{3N\lambda^4} \quad\text{.......................(18).}$$

This important equation was first deduced by Lord Rayleigh from the elastic solid theory of light, the scattering being supposed to be due to a weighting of the medium by the small particles. Rayleigh further showed that the same relation held in the electromagnetic theory if the scattering is due to local changes of inductive capacity, while the above investigation shows that it must generally hold whatever theory we adopt. But while there is a general connexion between the coefficient k and the refractive index, we cannot calculate the value of either of them without specifying the properties of the particles.

According to our present views of refraction and dispersion the molecules act as resonators, so that each of them may be considered to be the seat of an electric doublet, which becomes the source of the scattered light. If M be the moment of this doublet, the electric force at a distance r is, according to Art. 181, $2\pi\omega VM \sin\theta/r\lambda$. Giving to A and B the same meaning as in (86) Art. 151 and writing $e\xi$ for the moment of the doublet, the electric force in the scattered light is represented by

$$[A\cos(\omega t - lr) + B\sin(\omega t - lr)]\, e\omega^2 R_0 \sin\theta/4\pi r.$$

The comparison of this equation with (13) gives the relationship between the present A, B and the previous A_1, B_1. The equations (16) and (17) become by the substitution of the new coefficients:

$$2(\mu - 1) = eANV^2 \qquad (19),$$

$$k = e\omega BNV \qquad (20).$$

The retardation of phase at the source of the scattered light is B/A, which from (19) and (20) is equal to $kV/2\omega(\mu-1)$ or, in terms of the wave-length, to $k\lambda/4\pi(\mu-1)$.

Substituting the value of k from (18) we find the retardation of phase to be equal to $8\pi^2(\mu-1)/3N\lambda^3$. In the case of gases at normal pressure $N\lambda^3$ is of the order of a million for blue light, so that only a very high value of μ would make the change of phase appreciable. The values required can only be found in close contiguity to an absorption line in the region of selective refraction.

The elementary treatment of the problem which we have given may be replaced by the investigation of Art. 151. It will be noticed that equation (40) of that article becomes identical with (19) when κ is a small quantity, the square of which may be neglected, and when the refractive index does not differ much from unity. The constants κ and k are easily seen to be connected by $4\pi\kappa = k\lambda$.

If each molecule is supposed only to possess one electron having a free period $2\pi/n$ we may apply Art. 152 to determine the value of A. The light scattered per unit volume then becomes

$$k = \frac{e^2A^2N\omega^4}{6\pi} = \frac{8\pi Ne^4\omega^4}{3\rho^2(n^2-\omega^2)^2} = \frac{8\pi Ne^4\lambda_1^4}{3\rho^2(\lambda^2-\lambda_1^2)^2},$$

where λ_1 is the wave-length corresponding to the free period. When λ_1 is small compared with λ we are in the region of the ordinary dispersion of transparent media, and, as a first approximation, the scattering is inversely proportional to the fourth power of the wave-length, but it may become very considerable, as we have already seen, when λ becomes nearly equal to λ_1.

The first serious discussion of the colour of the sky based on the effect of small suspended particles is due to Lord Rayleigh* though the idea itself was not new at the time and had received support from Tyndall's experiments with precipitated clouds. Rayleigh showed that both the colour and the polarizing of the sky may be accounted for by the effect of scattering. He returned to the question subsequently† in a paper in which the question is raised whether the molecules of air are by themselves and without the help of suspended extraneous matter sufficient to explain the observed luminosity of the sky. The conclusion arrived at was: "that the light scattered from the molecules would suffice to give us a blue sky, not so very greatly darker than that actually enjoyed."

If atmospheric absorption were completely accounted for by scattering, the total light absorbed in the atmosphere might be calculated by means of the expression (18). If this is put into the form

$$k = \frac{32}{3}\pi^3\left(\frac{\mu-1}{N}\right)^2\frac{N}{\lambda^4}$$

and we remember that N is proportional to the density D and that in the case of gases $(\mu-1)/N$ is independent of pressure and temperature we see k is of the form cD where c only depends on the wave-length and on the nature of the gas. From $E^{-1}dE = kdx$ we therefore calculate that the fraction of light which is transmitted is equal to $e^{-\int cDdx}$. If H represent the height of the homogeneous atmosphere of normal density D_0, the fraction of light transmitted, when a pencil of light traverses the atmosphere vertically downwards, is e^{-kH} where we may substitute again for k its value of (13), putting for N the number of molecules per cubic centimetre when the gas is at normal pressure and temperature, and for μ the corresponding refractive index. The value $2{\cdot}72 \times 10^{19}$ obtained by Rutherford and Geiger for N agrees so well with other recent determinations that we may assume it to be substantially correct, and the theoretical transmission of the atmosphere may therefore be compared directly with Abbot's experimental results. Unfortunately Abbot does not give the height of the barometer on Mount Wilson which I have assumed to be 614 mm. corresponding with the average calculated value at an altitude of 1780 metres.

* *Collected Works*, Vol. I. p. 87.
† *Collected Works*, Vol. IV. p. 397.

In Table XVII. the first column gives the wave-length: the second column contains the observed values of the transmitted energy for Washington, taking all observations into account, while the third column gives the number calculated from the observations on Feb. 15, 1907, when the air seemed exceptionally clear. The calculated values are entered into the fourth column. The last three columns give the corresponding numbers relating to Mount Wilson. The selected clear day in this case was Oct. 11, 1906.

TABLE XVII.

Wave-length	Washington			Mount Wilson		
	Observed mean	Observed clear day	Calculated	Observed mean	Observed clear day	Calculated
4×10^{-5}	0·55	0·72	0·71	0·73	0·76	0·76
5	·70	·84	·87	·85	·89	·89
6	·76	·87	·94	·89	·92	·95
7	·84	·90	·96	·94	·96	·97
8	·87	·94	·98	·96	·99	·98
10	·90	·96	·99	·97	·99	·99

The close agreement between the two last columns shows that on a clear day on Mount Wilson atmospheric absorption is practically accounted for by the scattering of the molecules of air. There is a slight indication of selective absorption in the orange that may be due to aqueous vapour, but otherwise the agreement is good. It is remarkable that even at Washington the calculated absorption for the light should so nearly agree with the calculated value: this means that even at the sea level the greater part of the absorption on a clear day is due to scattering by the molecules of air. We also possess measurements of the luminosity of the sky made by Majorana near the crater of Mount Etna and on Mont Rosa. These are quoted and discussed by Lord Kelvin*, who took account of the illumination due to light reflected from the earth and that scattered by the atmosphere itself. If e and e' be the proportion of this secondary illumination on Mount Etna and Mont Rosa respectively and f, f' denote the light scattered by the molecules of air per unit volume, Kelvin expressed his result in the form:

$$N = \frac{2·50}{f(1-e)}\, 10^{19} = \frac{3·58}{f'(1-e')}\, 10^{19}.$$

Lord Kelvin uses these equations by introducing estimated values for e and e', and he concludes that the observations indicate the

* *Baltimore Lecture*, p. 307 seq.

probability that N is at least as large as 10^{20}. Our present knowledge does not admit so high value, and if we introduce the figure $2{\cdot}72\ 10^{19}$ already quoted, we see that e and f being fractions, Majorana's results on Mount Etna can only be reconciled with Lord Kelvin's calculations if e be practically zero and f nearly equal to unity. The measurements on Mont Rosa are not reconcilable at all. Considering that Majorana's method of observation consisted in comparing directly light from the sun with that reflected by different parts of the sky, it is perhaps surprising that they yield a value of N which is so nearly correct; but as regards the loss of light due to scattering, the method cannot compare in accuracy with that employed by the American observers.

The theoretical result embodied in equation (18) is subject to certain limitations which restrict its application to gases. This is shown by the factor $(\mu - 1)^2$ in place of $\mu^2 - 1$, or (for solids) more correctly still of $(\mu^2 - 1)/\mu^2 + 2$ (see Article 166). The subject has been discussed theoretically by Raman, and experimental investigations by Lord Rayleigh* have yielded important results bearing on the subject. If the laws of scattering as stated at the beginning of this article, were strictly true, the light sent out in a direction at right angles to the incident beam should be completely polarized in the plane containing the incident and scattered ray. But the theory on which the conclusion is founded assumes that the scattering particle is small, and has symmetrical properties in all directions. This first assumption must hold at any rate for gaseous body, but we have no independent reason for believing that there is no asymmetrical element in a molecular structure. The experimental test presents formidable difficulties which have been overcome by Lord Rayleigh. If the incident light is in a direction OX and the light scattered be examined in a direction OY, the ratio of intensities of light vibrating along OX and OY respectively was measured. This ratio should be zero if there is absolute spherical symmetry. Eight gases were examined, and it was found that in the case of Argon and Helium the fraction is small, but exceeded 15 °/。 in the case of Nitrous oxide. According to equation (18) the opacity should be proportional to $(\mu - 1)^2$. This was tested by comparing the transparencies of a gas with that of air. The inverse ratio of the two was found to be—within experimental errors—equal to the squares of the refrangibility which is measured by $\mu - 1$.

183. The pressure of light. A wave of light passing through any substance sets up an electric disturbance within the molecule, the effect of which, so far as scattering of energy is concerned, has been discussed in the previous article. We now enquire whether any electromagnetic forces are called into play by the mutual action of the

* *Proc. Roy. Soc.* A, Vol. xcv. pp. 155, 476; Vol. xcvii. p. 435; Vol. xcviii. p. 57.

electromagnetic field of the radiation and the electric current in the molecule. We take the incident light to be plane polarized and denote the electric force in the direction of z by R: the corresponding magnetic force acts in the direction of y, and by (13) Art. 136, is $\beta = -R/V$. The current in the molecule is $e\zeta$ where we may take as the most general form for ζ that given by (36) Chapter XI. This gives for the current:

$$e\zeta = \frac{e}{4\pi}\left(A\,\frac{dR}{dt} - \frac{B}{\omega}\,\frac{d^2R}{dt^2}\right)$$
$$= \frac{e}{4\pi}\left(A\,\frac{dR}{dt} + \omega BR\right).$$

The electromagnetic force acting on the molecule is equal to the product of the current and magnetic force. The term which is proportional to RdR/dt is periodic with an average value of zero, so that the term in B only need be considered. This gives for the average value of the force: $e\omega BR_0^2/8\pi V$ where R_0 is the maximum value of R_0. If there are N molecules per unit volume we must multiply by N to obtain the total force per unit volume. The factor B is intimately connected with the coefficient of extinction R as shown by equation (20) of the previous article, which gives for the force per unit volume: $kR_0^2/8\pi V^2$, or k times the energy per unit volume of the incident beam. This force is in the direction of wave propagation, and hence the result of the investigation is in full agreement with Maxwell's conclusion that light exerts a pressure equal to its energy per unit volume. The agreement is proved by considering a layer of unit surface and thickness dx. If the incident energy be E, the energy leaving the layer is, by definition of k, $E(1-kdx)$. The diminution of energy is $kEdx$ and the diminution per unit volume is kE; there is therefore an excess of pressure on the front of the layer which is k times the total light pressure. The importance of the above investigation lies in its proving that the light pressure acts in the same way when the body is gaseous as when it is solid or liquid, and that scattering is equally effective to absorption.

184. Group Velocity. In investigating the mechanics of a wave motion we are generally led to a Differential Equation and this is often satisfied by what we have called a "homogeneous" train of waves. If the velocity of propagation is then found to be independent of the wave-length, the propagation of plane waves can always be represented by an expression of the form $f(x \pm vt)$, so that a wave of any shape is propagated with a definite velocity without change of type. But this happens only in the simplest cases. When the wave velocity depends on the frequency, the wave-length alters in shape as it proceeds. Though the wave is not homogeneous in type, it can always be represented analytically as a superposition of homogeneous waves. As a simple case, take two such waves having slightly different lengths and rates of propagation, their displacements being $\cos l(Vt-x)$ and $\cos l'(V't-x)$, and their resultant

$$\cos l(Vt-x)+\cos l'(V't-x)$$

which is equal to

$$2\cos\left\{\frac{l'V'-lV}{2}t-\frac{l'-l}{2}x\right\}\cos\left\{\frac{l'V'+lV}{2}t-\frac{l'+l}{2}x\right\}\ldots(21).$$

The first factor passes through its period in a time $4\pi/(l'V'-lV)$ while the periodic time of the second is $4\pi/(l'V'+lV)$. If $l'-l$ and $V'-V$ be sufficiently small, it takes many periods of the second factor to produce an appreciable difference in the first factor. Hence we may say that the resultant effect is that of a wave having a length approximately equal to that of either train of waves, and an amplitude which varies slowly. The velocity of the waves in this group is $(l'V'+lV)/(l'+l)$ or to the first approximation, equal to that which corresponds to a wave-length $2\pi/l$. To find the velocity of the group we must fix our mind on some special feature which may be chosen to be the maximum amplitude. For $t=0$, this lies at the origin, and generally the amplitude has its maximum whenever

$$(l'V'-lV)t-(l'-l)x=0 \quad\ldots\ldots\ldots\ldots\ldots\ldots(22).$$

The highest point of the wave travels forward therefore at a rate which is $(l'V'-lV)/l'-l$ or, on the supposition of nearly equal values of l and l', we may write for the group velocity

$$U=\frac{dlV}{dl}=\frac{dV/\lambda}{d1/\lambda}$$

$$=V-\lambda\frac{dV}{d\lambda}\quad\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(23),$$

showing the dependence of the group velocity on the variation of the wave velocity with the wave-length.

The explanation of the propagation of groups was first given by Stokes while Osborne Reynolds pointed out its connexion with the propagation of energy. Equation (23) is due to Lord Rayleigh*. Professor Lamb† has given an instructive proof of this equation, in which the group is identified, not by a difference in amplitude, but by a difference in the distances between prominent points of the group. Let the group consist of waves approximately of the simply periodic character but with a gradual change in the distance from crest to crest. The group velocity will be the velocity with which a particular distance between two successive crests moves. The wave-length λ may here be considered to be a function of x and of t. The rate of change of λ at a

* *Collected Papers*, Vol. I. p. 322, *Sound*, Vol. I. § 191 and Appendix.

† *Proc. London Math. Soc.*, Sec. II. Vol. I. p. 473 (1904).

point which moves with velocity $\frac{dx}{dt}$ is by the rules of the differential calculus $\frac{\partial\lambda}{\partial t}+\frac{\partial\lambda}{\partial x}\frac{dx}{dt}$. If the velocity of the point is equal to the group velocity, then by the definition of U the wave-length is constant, hence

$$\frac{\partial\lambda}{\partial t}+U\frac{\partial\lambda}{\partial x}=0 \quad\text{..........................(24).}$$

Now let the point move with velocity V, *i.e.* follow one crest. The next crest will move with velocity $V+\lambda\frac{\partial V}{\partial x}$ or $V+\lambda\frac{\partial V}{\partial\lambda}\cdot\frac{\partial\lambda}{\partial x}$, hence $\lambda\frac{dV}{d\lambda}\cdot\frac{d\lambda}{dx}$ measures the rate at which the wave-length increases. This gives

$$\frac{\partial\lambda}{\partial t}+V\frac{\partial\lambda}{\partial x}=\lambda\frac{\partial V}{\partial\lambda}\cdot\frac{\partial\lambda}{\partial x} \quad\text{........................(25).}$$

By combining (24) and (25), we return to equation (23).

One word of caution may be necessary; when we speak of the velocity of a group, we do not mean to imply that the whole of the group moves forward without any alteration just as if it were a single wave. The first variable factor of (21) always has its maximum value when the condition (22) is satisfied, but the second factor continuously changes and at intervals of time which are equal to half a period, this factor is alternatively ± 1, so that the maximum amplitude after such an interval is converted into a minimum. The essential point is, that at periodically recurring intervals, the group regains its original feature, and the distance through which the group has moved forward divided by the interval is called the velocity of the group.

The following Table given by Rayleigh is interesting as giving the relation between group and wave velocities in particular cases.

$V\propto\lambda$,	$U=0$.	Reynolds' disconnected pendulums.
$V\propto\lambda^{\frac{1}{2}}$,	$U=\frac{1}{2}V$.	Deep water gravity waves.
$V\propto\lambda^{\circ}$,	$U=V$.	Aerial waves etc.
$V\propto\lambda^{-\frac{1}{2}}$,	$U=\frac{3}{2}V$.	Capillary water waves.
$V\propto\lambda^{-1}$,	$U=2V$.	Flexural waves in elastic rods or plates.

The last two examples show that it is possible for the group to travel more quickly than the individual wave.

When the law connecting V and λ is $V=a+b\lambda$, the group velocity is independent of λ and the variations in the shape of the group may in special cases be followed out in detail*. Fig 181 represents the successive stages of a group, the shape of which is represented by the equation

$$y = \frac{h^2}{h^2 + x^2}.$$

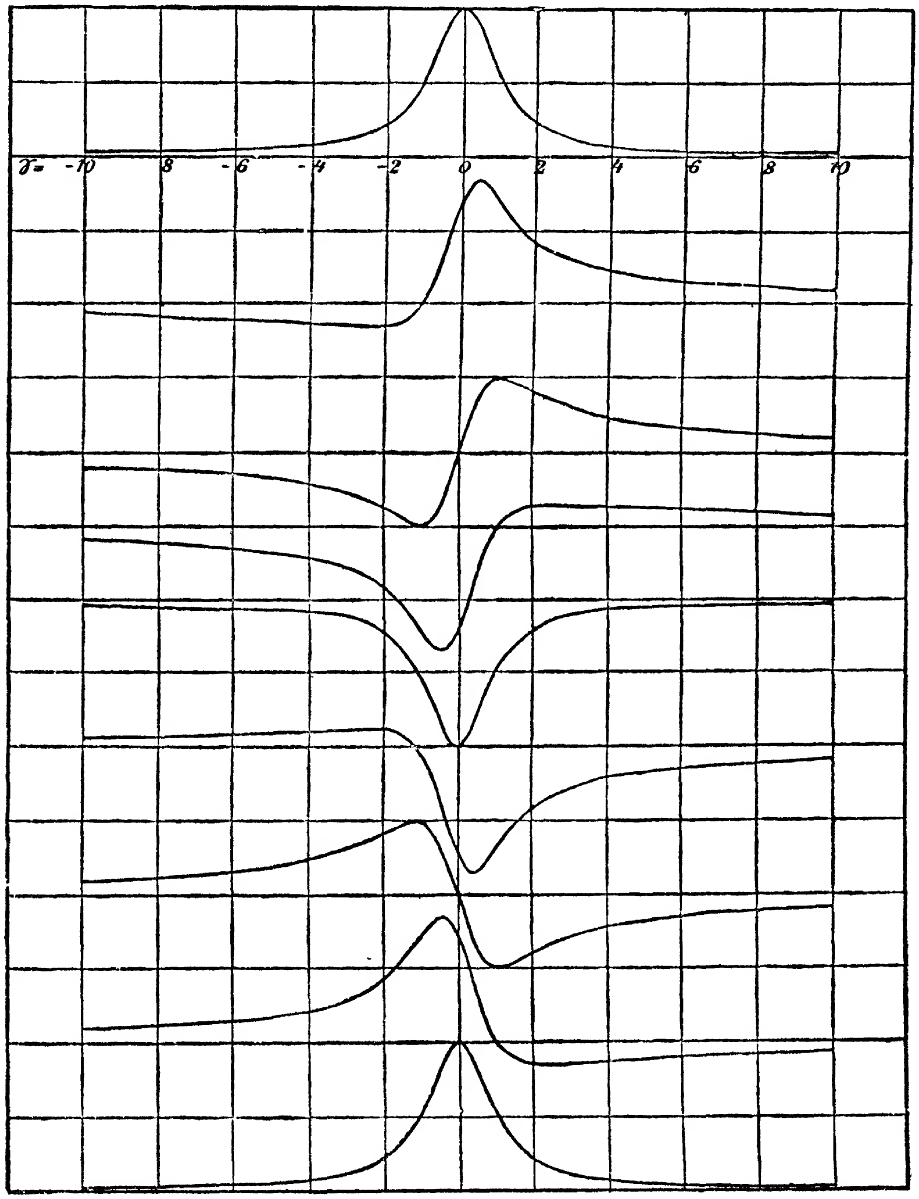


Fig. 178.

An interesting question arises in the case of the propagation of light within an absorption band. As explained in Art. 157 the wave velocity may increase with diminishing wave-length. In that case let V_0 and λ_0 represent the velocity and wave-length in vacuo, and

* Schuster, *Boltzmann*, *Festschrift*, p. 569.

let $dV/d\lambda_0$ be negative. As $V\lambda_0 = V_0\lambda$, we obtain by differentiation with respect to V

$$\lambda_0 + V\frac{d\lambda_0}{dV} = V_0\frac{d\lambda}{dV},$$

or

$$\lambda + \frac{V^2}{V_0}\frac{d\lambda_0}{dV} = V\frac{d\lambda}{dV}.$$

As the second term on the left-hand side is negative, it follows that $\lambda\frac{dV}{d\lambda} > \lambda$, which shows that the group velocity is in the opposite direction to the wave velocity. If there is a convection of energy forward, the waves must therefore move backwards. In all optical media where the direction of the dispersion is reversed, there is a very powerful absorption, so that only thicknesses of the absorbing medium can be used which are smaller than a wave-length of light. Under these circumstances it is doubtful how far the above results have any application. But Professor Lamb* has devised mechanical arrangements in which without absorption there is a negative wave

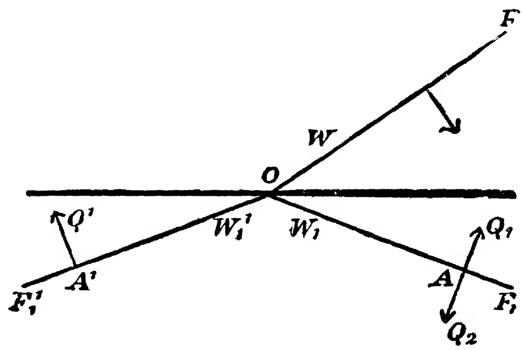


Fig. 179.

velocity. One curious result follows: the deviation of the wave on entering such a medium is greater than the angle of incidence, so that the wave normal is bent over to the other side of the normal as indicated in Fig. 179. This is seen at once by considering that the traces on the refracting surface of WF and W_1F_1, the incident and refracted wave-fronts, must move together. If we were to draw the wave-front in the usual way parallel to $W_1'F_1'$ and the waves moved backwards in the direction $A'Q'$, the intersection O of the refracted wave and surface would move to the left, while the intersection of the incident wave moved to the right. By drawing the refracted wave-front in the direction W_1F_1 the required condition can be secured. The individual waves move in the direction AQ_1 but the group moves in the direction AQ_2.

* *Proceedings London Math. Soc.* Sec. II. Vol. I. p. 473 (1904).

CHAPTER XIV.

FURTHER DISCUSSION OF THE NATURE OF LIGHT AND ITS PROPAGATION.

185. Preliminary Remarks. Light enters into our consciousness through the effects on our sensitive organs whose powers may be increased by suitable experimental appliances. When we speak of the "nature" of light we try to form some mental picture of what constitutes light before it has entered our spectroscopes or other optical instruments, and it becomes necessary therefore to examine what modifications light undergoes in passing through such instruments. If—as an example—we were to look upon a spectroscope as an appliance capable of analysing white light, in the manner a chemist analyses a compound body by separating the constituents it contains, we might be led to believe that the highly homogeneous radiations which leave the spectroscope have a real existence in the light that entered it. This —as the late Lord Rayleigh pointed out—is an error: it is the spectroscope that converts the white light into homogeneous radiations. Having satisfied ourselves with regard to instrumental effects, we have to consider the ultimate receiving screen such as the retina or the photographic plate. How much our judgment is affected by the peculiarities of these receivers may be recognized if we try to imagine how radically our impressions would be altered if our eyes were equally sensitive to radiations of all kinds, so as to give us simply a measure of their intensities*. What is true of instrumental analysis is equally true of its mathematical treatment. The process of the treatment may affect our conclusions.

186. Application of Fourier's theorem. Gouy's treatment. This theorem gives us the most powerful mathematical method of treating variable functions, that without necessarily being periodic oscillate between finite limits. We begin by considering in greater detail the series that has already been mentioned in Art. 10. We consider a ray of plane polarized light and fix our attention on a point P over which the disturbance passes. If the velocity at P be v, we may, in the most general case, express it as a function of the time, $f(t)$. Let us follow the motion from a time $t=0$, to a time $t=T$. According to Fourier's theorem, which has already been explained in Art. 10, we may write

$$f(t)=a_0+a_1\cos(2\pi t/T)+a_2\cos(4\pi t/T)+a_3\cos(6\pi t/T)$$
$$+b_1\sin(2\pi t/T)+b_2\sin(4\pi/T)+b_3\sin(6\pi t/T) \quad \ldots\ldots(1).$$

Assuming that it is always possible to express v in terms of such a series, we may easily determine the value of any coefficient a_s by multiplying

* Schuster, *Phil. Mag.* Vol. XXXVII. (1894).

22—2

both sides by $\cos(2\pi st/T)$ and integrating between the limits 0 and T. It will be found that on the right-hand side all integrals have the same value at both limits except that one which has a_s for coefficient.

We find similarly any coefficient b_s by multiplying both sides by $\sin(2\pi st/T)$ and integrating between the same limits.

We thus obtain
$$a_0 = \frac{1}{T}\int_0^T f(\tau)\,d\tau \qquad\qquad (2),$$

and for the other coefficients,
$$\left.\begin{aligned} a_s &= \frac{2}{T}\int_0^T f(\tau)\cos(2s\pi\tau/T)\,d\tau \\ b_s &= \frac{2}{T}\int_0^T f(\tau)\sin(2s\pi\tau/T)\,d\tau \end{aligned}\right\} \qquad\qquad (3),$$

where the variable has been altered for convenience in future use from t to τ.

As a_0 expresses the difference between the displacements of the point P at times $t = T$ and $t = 0$, we may, in the case of periodic motions, by choosing the time T to be very large, make a_0 as small as we like. We shall therefore neglect this quantity. The remainder of the series may then be written
$$v = v_1 + v_2 + \ldots v_s + \qquad\qquad (4),$$
where
$$v_s = r_s \cos\{(2\pi st/T) + \theta_s\} \qquad\qquad (5),$$
r_s and θ_s being two quantities which may be determined in the usual way from a_s and b_s.

Each term of the series (1) is identical in its analytical expression with what we have called a simple periodic motion giving rise to a homogeneous wave, but we must bear in mind that the equation only holds during a certain time interval, and that homogeneous light necessarily implies an infinite succession of waves. Hence some care is necessary in the application of the formula. We may however, as we are at liberty to choose the time T as large as we like, express the whole disturbance as being formed by the superposition of a number of disturbances each of which may be made as nearly identical as we please with homogeneous light.

In the analytical discussion of diffraction and refraction, we have based our investigation on the treatment of homogeneous waves, and where the light was not homogeneous, we have assumed that the total effect as regards intensity, could be represented as being equal to the sum of the separate effects of a large number of homogeneous vibrations. This requires justification. Imagine the disturbance, which may be of quite arbitrary type, to pass through any optical system and confine the attention to that point of the system where the observations are carried out. When T is very large, we may, except possibly near the limits of time, treat each term of the series (1) as being due to a homogeneous wave. At the point considered, the velocity may be expressed as the sum of terms of the form (5), with altered values of r and θ.

Calculate now the average square of the velocity during the interval T. The square of the right-hand side of (4) contains products such as $v_n v_s$ and hence the expression for the average value of v^2 contains terms of the form

$$\frac{r_n r_s}{T}\int_0^T \cos\{(2\pi nt/T)+\theta_n\}\cos\{(2\pi st/T)+\theta_s\}\,dt.$$

n and s being integers, the integral is easily shown to be zero. The remaining terms to be considered are of the form

$$\frac{r_s^2}{T}\int_0^T \cos^2\{(2\pi nt/T)+\theta_s\}\,dt=\tfrac{1}{2}r_s^2,$$

and hence for the average value of v^2 we find $\frac{1}{2}\Sigma r^2$; but this is exactly the same expression we should have found, if we had treated each component of the series (4) as an independent homogeneous vibration. The intensity of the luminous disturbance at any time is proportional to v^2, and our proof of the independence of the separate vibrations as regards energy only applies to the average energy extended over a very long range of time. The relevancy of the proposition as regards light depends on the fact that in our optical investigations we may treat the sources of light to be constant, so that the average energy is independent of the length of the time interval. This important remark was first made by Gouy*, to whom the whole of the above investigation is due. The simplification in the treatment of non-homogeneous light which was first made at the end of Art. 20 now finds its complete justification, and we are at liberty, whenever it is convenient, to represent white light by superposing a number of homogeneous vibrations having periods which lie very close together. But we are equally at liberty to assume any other representation so long as its resolution by Fourier's theorem gives us a distribution of intensity equal to that of the observed one. Gouy pointed out that we can regard white light as being made up of a succession of perfectly irregular impulses. The type of the impulse is quite arbitrary so long as the conditions as regards distribution of intensity are satisfied.

187. Application of Fourier's integral. Lord Rayleigh's investigations. Lord Rayleigh† had independently arrived at conclusions similar to those of Gouy, and has more definitely investigated the type of impulse, an aggregation of which may be considered to constitute white light‡.

* *Journal de Physique*, Vol. v. p. 354 (1886).

† *Collected Works*, Vol. III. p. 60.

‡ *Collected Works*, Vol. III. p. 268.

If we write (1) in the form

$$v = \sum_{s=0}^{s=\infty} \{a_s \cos(2\pi st/T) + b_s \sin(2\pi st/T)\},$$

and substitute the values of a_s and b_s from (2) and (3), we obtain

$$f(t) = \frac{2}{T}\left\{\frac{1}{2}\int_0^T f(\tau)\,d\tau + \sum_{s=1}^{s=0}\int_0^T f(\tau)\cos\{2\pi s(\tau - t)/T\}\right\}d\tau.$$

If now T is allowed to increase indefinitely, we may write $\omega = 2\pi s/T$ and for the increase of ω in successive terms of the sum, $d\omega = 2\pi/T$. The first term on the right-hand side vanishes. By substituting integration for summation, we obtain Fourier's theorem in the form

$$\pi f(t) = \int_0^\infty d\omega \int_0^T f(\tau)\cos\omega(\tau - t)\,d\tau \quad\ldots\ldots\ldots\ldots\ldots(6).$$

This equation represents the way in which any given function $f(t)$ may be analysed into its homogeneous components The next step is to find how much energy is to be ascribed to each small range of periods defined by the values of ω. This is most easily done by means of a theorem expressed by the following equation*:

$$\pi\int_0^\infty f(t)\,\phi(t)\,dt = \int_0^\infty (A_1A_2 + B_1B_2)\,d\omega,$$

if

$$A_1 = \int_0^{+\infty} f(\tau)\cos\omega\tau\,d\tau; \qquad B_1 = \int_0^{+\infty} f(\tau)\sin\omega\tau\,d\tau,$$

$$A_2 = \int_0^{+\infty} \phi(\tau)\cos\omega\tau\,d\tau; \qquad B_2 = \int_0^{+\infty} \phi(\tau)\sin\omega\tau\,d\tau.$$

If $f(t)$ expresses a vector, the square of which is proportional to the energy, $\int_{-\infty}^{+\infty}[f(t)]^2\,dt$ may be taken as the measure of the total energy of the disturbance, and by the above theorem,

$$\pi\int_{-\infty}^{+\infty}[f(t)]^2\,dt = \int_0^\infty (A^2 + B^2)\,d\omega \quad\ldots\ldots\ldots\ldots\ldots(7),$$

where

$$A = \int_{-\infty}^{+\infty} f(t)\cos\omega\tau\,d\tau; \quad B = \int_{-\infty}^{+\infty} f(\tau)\sin\omega\tau\,d\tau.$$

It follows that $(A^2 + B^2)/\pi$ may be taken as the measure of the energy in the range defined by $d\omega$, the frequency being $\omega/2\pi$.

As an example, Lord Rayleigh takes a disturbance originating at a point and having at any time a velocity given by

$$f(t) = e^{-c^2t^2} \quad\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(8).$$

* Schuster, *Phil. Mag.*, Vol. XXXVII. p. 509 (1894).

This disturbance is very small when t is large on the negative or positive side, and has a maximum for $t=0$. In this case

$$A = \int_{-\infty}^{+\infty} e^{-c^2t^2} \cos \omega\tau\, d\tau = \frac{\sqrt{\pi}}{c} e^{-\omega^2/4c^2},$$

$$B = \int_{-\infty}^{+\infty} e^{-c^2\tau^2} \sin \omega\tau\, d\tau = 0.$$

Hence the energy included in the range between ω and $\omega + d\omega$ is

$$c^{-2} e^{-\omega^2/2c^2} d\omega.$$

This represents a distribution of intensity resembling to some extent that observed in the case of the light emitted from black bodies. The example is sufficient to show that it is possible to represent white light as being due to the emission of a succession of disturbances, each of which roughly resembles that represented by (8). The larger the value of c, the more sudden will each disturbance be, approaching ultimately to an impulsive motion.

188. White light analysed by grating. It is interesting to follow out the effect of a grating in modifying a disturbance of any shape. For this purpose we must define the action of a grating a little more closely. Let s be measured along the grating, at right angles to its lines, and $f(s, t)$ be the displacement. The grating modifies the disturbance in a periodic manner, and we obtain the simplest kind of modification by assuming that the disturbance in the reflected light is equal to $\cos qs \,.\, f(s, t)$, $2\pi/q$ measuring the distance between the lines of the grating. An imaginary grating having this property was made use of by Lord Rayleigh in his article on the wave theory in the *Encyclopaedia Britannica*. It may conveniently be called a simple grating, and it can be shown that all real gratings may be represented by the superposition of a number of simple gratings.

It can be shown* that if a disturbance, originally coming from a point, is spread over a plane wave-front at right angles to x with a velocity determined by $\psi(Vt - x)$ and falls on a simple grating, the displacement in the reflected beam is determined by the equation

$$2\pi\phi(t) = \frac{h \cos \beta}{VF} \int_{-l}^{+l} \cos qs\, \psi(Vt - \gamma s)\, ds \quad \text{.........(9).}$$

The displacement is measured at the focus of the lens collecting it The other quantities which occur in the equation are defined as follows: h = height of grating; $\gamma = s(\sin\beta - \sin\alpha)$; α = angle of incidence; β = direction of reflected beam; $2l$ = width of grating.

In Fig. 180 let the thick line represent the velocity of the disturbance travelling in the positive direction, and the thin line

* Schuster, *Phil. Mag.* Vol. xxxvii. p. 509 (1894)

the curve $y = \cos qx$, drawn from the point $x = 0$ to the point $x = 2\pi N/q$. Then if y' be the ordinate of the thick curve, the displacement of the focus of the telescope after reflexion from the

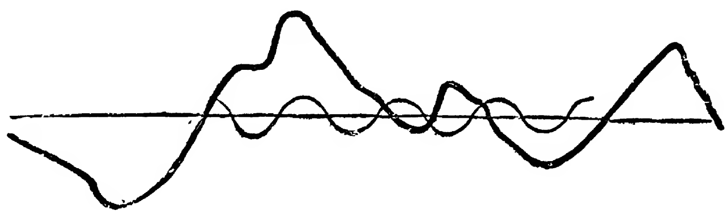


Fig. 183.

grating at a certain time, is seen by (9) to be proportional to $\int yy'dx$. The displacement at all times is obtained by letting the wave travel forward, the cosine curve remaining in the same position. The formula (9) and the geometrical interpretation just given, bring out clearly the analogy between the action of the grating and the integration involved in the calculation of the coefficients in Fourier's series. I must refer to the original paper for a detailed discussion of the application of the above equation, but the following two special cases may help the student in clearing his ideas.

Case I. The incident beam is homogeneous. The light reflected from the grating in any direction is then also homogeneous, and has the same period as that of the original light. That is to say, whatever the periodicity of the grating, it has no power to alter the periodicity of the disturbance. The distribution of amplitude in different directions is the same as that which has already been obtained in the Chapter on Gratings.

Case II. The incident disturbance consists of a single impulsive velocity. The disturbance at the focus of the telescope consists then of an impulsive displacement followed by a vibration represented by a cosine curve, continuing for as many periods as there are lines in the grating. The period is the same as that of the homogeneous vibration which, having the wave-front parallel to that of the disturbance considered, would have its principal maximum at the focus of the telescope; that is to say, the periodicity in the reflected beam depends now on the direction in which the telescope points.

189. White light analysed by dispersive media. The mechanism by means of which a grating converts an impulsive motion into a regular succession, with one predominant period, is easily explained. The action of a prism is a little more difficult to understand. Nevertheless, we know that a prism behaves in the same way as a grating, and the method of its action suggests itself as soon as we consider that the separating power of the prism is due to the unequal velocity of different wave-lengths through its substance. An impulsive motion therefore, started in a dispersive medium, cannot remain an impulsive motion, but the disturbance lengthens out as the wave proceeds. Having again recourse to Fourier's theorem, we may analyse an impulse, and imagine it to be made up of a number of different groups of waves of different lengths. These groups, according to Art. 184, are propagated with different velocities so that a separation of the different wave-lengths takes place, except in the particular case considered in that article, in which the law of propagation is such that the group velocity is the same for all wave-lengths. This being the simplest conceivable law, we may consider the action of a prism made of a substance for which the group velocity is constant. A plane impulsive motion reaching such a prism obliquely is refracted, and we can draw a plane over which the disturbance is spread in the prism in the ordinary way, by substituting the group velocity for the wave velocity. The plane of the disturbance stands therefore oblique to what, in ordinary refraction, would be the wave-front, and the optical distance from the original wave to the different points of the plane of disturbance is not the same.

It has also been pointed out that as the group proceeds, maxima are periodically changed into minima, and vice versa. On the plane of disturbance therefore the motion will not be everywhere in the same direction, but will change alternately from one direction to the other. If we now follow this plane of disturbance proceeding with the group velocity as it is refracted out of the prism, we obtain another plane of disturbance oblique again to what, under ordinary circumstances, would be the wave-front. Its position would be the same as that of an ordinary wave-front which has passed through the prism with the group velocity. If the emergent wave be now received by a lens, the disturbance at the focus of the lens consists of a periodic motion, the different parts of the plane of the disturbance passing through the focus at different times. The greater the resolving power of the prism, the greater will be the number of inversions in the plane of disturbance after it has left the prism and therefore the more will the light passing through the focus of the lens be homogeneous. In this way we may convince ourselves that the action of a prism is identical with that of a grating. Although our reasoning is strictly correct only for a prism composed of a material which has a definite law of dispersion, the result must be the same in other cases, because, fixing our attention on a certain narrow range of wave-lengths, we may always consider the groups of waves within this range to proceed through the material with constant velocity. We may therefore apply the above reasoning to each such narrow range separately.

190. Doppler's Principle. If the spectra of stars are carefully examined it is found that in many cases the absorption lines when compared with the emission lines of chemical elements are slightly displaced. This can be explained by the motion of the star in the line of sight. If a star receding from the earth with velocity v emits or absorbs light of period t_0, it will have moved through a distance vt_0 while the vibrating corpuscle is performing a complete oscillation. The light reaching the earth has to pass through this additional distance, and V being the velocity of light will take a time vt_0/V to do it. The period as it appears to a terrestrial observer will therefore be

$$t = \frac{V+v}{V} t_0.$$

If the star moves towards the earth we must give a negative sign to v. As the same reasoning can be applied to the motion of the earth towards or away from the star, the displacements observed furnish a measure of the relative velocity of star and earth in the line of sight.

191. Homogeneous Radiations. Radiations which exhibit a high degree of homogeneity only appear when the radiating body is in a gaseous state and must therefore be considered as emanating from the individual atoms. It was first pointed out by Lippich* and subsequently by the late Lord Rayleigh† that the dynamical theory of gases combined with Doppler's principle sets a limit to the homogeneity. A molecule, which if stationary emits a radiation of frequency N, will, if moving in the line of sight with velocity v relative to the observer, appear to emit a radiation of frequency n where

$$n = N(1 \pm v/V),$$

V being the velocity of light, the upper sign being taken if the molecule recedes from the observer. If the molecules move with the same velocity but indiscriminately in all directions, it may be shown that the radiation received will cover uniformly a range of frequency between the limits $N(1 \pm v/V)$. The translatory velocity of the molecule being small compared with the velocity of light the result will be that the homogeneous radiation will only be slightly affected and appear as a narrow band in the spectrum.

In a more detailed calculation Lord Rayleigh abandons the simplifying assumption that all molecules move with the same velocity and assumes Maxwell's law of distribution. In order to conduct the investigation so as to lead to results capable of being verified experimentally, he calculates the distance at which the widening just prevents the resolution of two closely adjacent similar lines, or, what comes to the

* *Pogg. Ann.* XXXVI. p. 466 (1889).

† *Collected Works*, Vol. I. p. 183.

same thing, the greatest difference in path at which interference between two pencils of light is still observable. He finds that this difference in path expressed in wave-lengths is 690 V/v. Fabry and Perot* were able in the case of the green mercury line still to observe interference with a difference of path of 790000 wave-lengths. If the molecular velocity is the sole cause of the widening of the lines we may use Lord Rayleigh's expression to calculate v which is then found to be equal to 16900. Assuming thermal equilibrium this velocity corresponds to 376° C. We are ignorant of the effects of temperature in highly rarefied gases rendered luminous by electric discharges, but the calculation tends to show that in the case of metallic vapours giving nearly homogeneous radiations, the widening of the bands, which sets a limit to their powers of interference, must in great part, if not entirely, be ascribed to the Doppler effect. No account is taken in the above reasoning of molecular impact, which, in the example chosen, is not likely to have an appreciable effect.

One further point should be noticed. In the case of white light it has been shown that the homogeneous radiation which emerges from the analysing spectroscope is an instrumental effect. It has no independent existence prior to its entry into it. When on the contrary we can obtain interference from a beam of light one part of which has been retarded relatively to the other, the regularity must exist in the beam itself. No theory of light can be said to be complete if it is inconsistent with the almost perfect regularity of the homogeneous vibrations sent out by the atoms, a regularity equalling that of a clock which keeps time within a second in twenty days. As the light we receive is the combined effect of a large number of luminous centres, there is little room for any variations in the fundamental constants which determine the period. As expressed by Clerk Maxwell: "Each molecule therefore, throughout the universe, bears impressed on it the stamp of a metric system as distinctly as does the metre of the Archives at Paris, or the double royal cubit of the temple at Karnac."

192. Series Spectra. Results of fundamental importance bearing on the mechanism of radiation have been derived from the study of the regularities observed in the frequencies of the quasi-homogeneous light emitted by gaseous atoms. These will be described in the concluding Chapters of this treatise, and we shall confine ourselves here to a short account of the principal experimental facts on which the more recent theoretical investigations are based. We shall identify homogeneous radiation by its wave-number (n) which represents the number of waves per unit length so that $n\lambda = 1$.

* *C.R.* Vol. CXXVIII. p. 1223 (1899).

Applying the connexion between the wave-length λ and the frequency f we can also write $n = V^{-1}f$ where V is the velocity of light. A considerable number of elements—more especially in the spectra of flames or weak spark discharges—shew certain groups of lines which together form a series in which the lines lie nearer and nearer together as we pass from the red to the violet. The typical case is that of the ordinary spectrum of hydrogen for which

$$n_m = N\left(\frac{1}{4} - \frac{1}{m^2}\right) \text{......}$$

N is a constant equal to 109678 and m an integer number having for successive lines the values 3, 4, etc. It will be seen that the number of lines is infinite and that as m increases they approach a definite limit $n = \frac{1}{4}N$, which may be called the convergence number. Thirty-four lines of this series have been identified, the first, for which $m = 3$, giving $\lambda = n^{-1} = 6562{\cdot}8$ which is the well-known red line of hydrogen. The series was first given in a different form by Balmer and is always connected with his name. Three series presenting certain characteristics are generally associated together. These were in many cases discovered and studied by Kayser and Runge who represented the wave-numbers by a series having three constants:

$$n_m = A - \frac{B}{m^2} - \frac{C}{m^4},$$

where m again is a number which takes all integer values. The series now generally adopted, which was first used by Rydberg in his important contributions to the subject, has the form:

$$n_m = A - \frac{N}{(m+\mu)^2}.$$

The constants here are A, N and μ, but as the value of N is found to have nearly equal values for different elements, Rydberg treats it as a universal constant, sacrificing to some extent the accuracy of the agreement between the formula and the observed wave-numbers in favour of simplicity of expression.

The three series which as stated are generally associated may be written conveniently:

$$P(m) = P\infty - N/(m+P)^2, \qquad m = 1, 2, 3 \ldots.$$
$$S(m) = S\infty - N/(m+S)^2, \qquad m = 2, 3, 4 \ldots.$$
$$D(m) = D\infty - N/(m+D)^2, \qquad m = 2, 3, 4 \ldots.$$

The symbol ∞ indicates that the wave-number $P(m)$ applies to the case $m = \infty$ so that $P\infty$, $S\infty$, $D\infty$ represent the convergence numbers. The first series is called the Principal Series, the second consisting of slightly broadened lines is called the Diffuse Series, the third is known as the Sharp Series. The general appearance of the distribution of lines in the three series is illustrated in Fig. 181 in which the wave-numbers increase from right to left. The figure shows that the Diffuse and Sharp Series converge to the same point. This is a property which holds

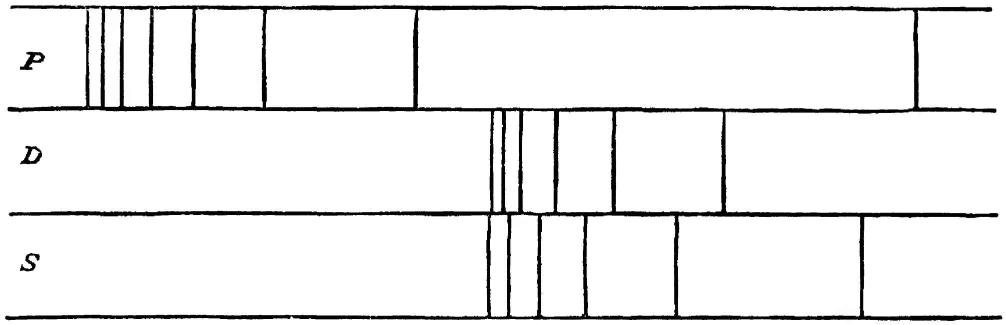


Fig. 181.

generally and we may express it by writing $S\infty = D\infty$. It is also found that the difference between the convergence number of the principal series and the common limit of the S and D series is very nearly equal to the wave-number of the principal series.

This law may be expressed by

$$P\infty - S\infty = P\infty - N/(1+P)^2,$$

or

$$S\infty = N(1+P)^2.$$

With a less degree of accuracy, Rydberg puts

$$P\infty = N(1+S)^2.$$

Using this value for the convergence numbers, he adopted the following symmetrical expressions for the three series:

$$P(m) = \frac{N}{(1+S)^2} - \frac{N}{(m+P)^2},$$

$$S(m) = \frac{N}{(1+P)^2} - \frac{N}{(m+S)^2},$$

$$D(m) = \frac{N}{(1+P)^2} - \frac{N}{(m+D)^2}.$$

These formulae are in general only approximate, but as we shall show subsequently they have a theoretical significance. The above account is mainly derived from Professor Fowler's Report on "Series in Line Spectra" to which the reader is referred for the numerical data and other details.

193. Infrared and Ultraviolet Radiations. The science of Optics taken in its literal sense should deal only with those radiations which affect our eyes. This would confine it to a narrow range extending over less than an octave and the limitation would be entirely unscientific as it would introduce an arbitrary physiological element. The true measure of the electromagnetic radiations, to some of which our eyes are sensitive, is their intensity measured by their thermal effects. For the sake of logical sequence we must risk a philological inexactitude and include in the theory of Optics all such radiations as have a molecular or atomic origin. From an experimental point of view the chemical effects which manifest themselves by their photographic action give us the widest range, covering at least a million octaves. The common nomenclature based on ocular observations conveniently divides the entire range into three parts, of which the so-called infrared, including wave-lengths longer than the visible ones, and the so-called ultraviolet, covering wave-lengths shorter than the visible ones, lie at the two sides of the region which directly affects our eyes. In the infrared Abney has obtained photographic effects as far as 27000 Å though the exact figure is doubtful, as the wave-length was determined by extrapolation. The observations of Rubens and his collaborators reaching as far as ·006 cm., *i.e.* 6×10^5 Å, have been referred to in Art. 153. In the ultraviolet, which may be said to begin at 3990 Å, we cannot go far beyond the visible part with glass prisms and lenses owing to their want of transparency. Quartz takes us a good step further, and Stokes with the help of a fluorescent screen was able to extend the observations to 1850 Å. Unless gratings are employed we are then obliged to use fluorspar as the refracting medium. But here another difficulty presents itself, for the air through which the radiations pass becomes strongly absorbent and the observations must be carried out in an atmosphere of hydrogen or *in vacuo.* V. Schumann to whom we owe the extension of our spectroscopic method in this direction was able to observe wave-lengths which he estimated to be about 1000 Å.

Ordinary photographic plates cannot be used in these experiments because the gelatine absorbs the light and prevents it from acting on the silver salts, and it required considerable experimental skill to devise a method for preparing sensitive films without the introduction of an absorbent medium. A considerable step in advance was made by Lyman* who substituted gratings for the prisms. Wave-lengths could then be determined with accuracy, and the spectra of a number of substances were mapped. More recently still Millikan† made further progress chiefly with the help of a specially constructed grating with suitable ruling. He could identify lines of nickel having a wave-length of 202 Å.

It is now generally accepted that the X-ray radiation discovered by Roentgen consists of ultraviolet waves. Some difficulty was felt at first in accepting this explanation because the radiation traversed substances without being refracted. This is, however, a direct consequence of

* *Astrophysical Journal,* Vol. XIX. p. 263 (1904).

† *Ibid.* Vol. LII. p. 47 (1920).

Sellmeyer's equation*, if the period be short compared with that of the free vibrations of the atoms.

Gradually evidence accumulated that the Roentgen rays were electromagnetic waves distinguished from ordinary light only by short wave-lengths. In a series of important investigations, summarized in his Bakerian Lecture†, Barkla proved that the rays could be polarized by scattering, and showed that the experimental facts he had obtained were inexplicable except on the transverse wave theory. Attempts were made by Haga and Wind to determine the wave-lengths by tracing the spreading out of a pencil of rays after passage through a narrow slit. They estimated the order of magnitude to be 10^{-8} cm. which subsequently proved to be correct, but the experiments were open to criticism and not generally accepted. Real progress was only made when, on Laue's ‡ suggestion that the regular spacing of atoms in crystals might serve to act as a three dimensional grating, Friedrich and Knipping, working in conjunction with Laue, proved the feasibility of the proposal. Their results were discussed and explained by W. L. Bragg§ and accurate experimental methods for more exact investigations were introduced by W. H. and W. L. Bragg‖. We cannot enter here into the important bearing of these researches on the theory of atomic structure in crystals. It must suffice to mention that the wave-lengths of the characteristic radiations of platinum, nickel and tungsten were found to have wave-lengths of 1·10, 1·66, and 1·25 10^{-8} cm. respectively. The measurements were subsequently extended to other elements by Moseley, who introduced a photographic method. The subject will be further referred to in Chapter XV.

194. The aberration of Light. Bradley who subsequently became Astronomer Royal discovered in the year 1726 that fixed stars describe in the course of a year elliptic curves, and he was able to explain this as being due to the earth's motion. The explanation is simple if we adopt the corpuscular theory of light. Consider a screen AB (Fig. 182) with a narrow aperture at O, and a similar screen PQ with a corresponding aperture at C. If a particle projected from the star moves in the direction OC, it should, after passing through O, ultimately pass through C. But if during the passage from O to C the earth moves in the direction PQ the path of the particle

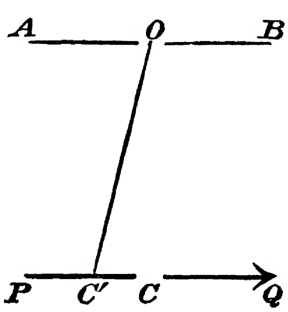


Fig. 182.

* *Nature*, Vol. LIII. p. 268 (1896).
† *Phil. Trans.* (A) Vol. CCXVII. p. 315 (1917).
‡ Drude's *Annalen*, Vol. XLI. p. 971 (1913).
§ *Proc. Cam. Phil. Soc.* Vol. XVII. p. 43 (1912).
‖ *Proc. Roy. Soc.* (A), Vol. LXXXVIII. p. 428 (1913).

relative to the earth will be along some line OC' slightly inclined to OC, so that the screen PQ would have to be shifted to the left through the distance CC' in order that the light be visible to an eye placed behind the aperture. The apparent line of sight is therefore OC' and its angular displacement is measured by v/V, if v be the component of the earth's motion which is perpendicular to the line of sight and V the velocity of light. A similar argument holds good in the undulatory theory. Let AB now represent the lens of a telescope. A wave-front after passing through the lens will be a sphere having C as centre where C must be considered to be fixed in space. But by the time the wave-front reaches the focal plane some other point C' moving with the earth will be in the position previously occupied by C, so that the apparent line of sight will, as before, be $C'O$.

195. Passage of light through moving bodies. There is one formidable difficulty which the undulatory theory of light with its doctrine of an æther permeating all bodies has to contend with. Is this æther fixed in space and unaffected by rapidly moving matter passing through it? It is natural that this question should have been raised at an early stage when the theory still stood on its trial. The explanation of the aberration of light given in the preceding article involves the assumption of a stationary æther, but if this æther partakes of the properties of ordinary solids and liquids it is difficult to believe that huge bodies such as the earth can sweep through it without carrying it partly or wholly along with them. Our theory of the aberration of light then fails, and on the basis of what has been called "classical dynamics" no satisfactory alternative seems possible. From the experimental side Arago was the first to make an attempt at attacking the problem. He observed the position of a star, the light passing through an achromatized prism before it entered the telescope. His object was to ascertain whether the deviation caused by the prism depended on the motion of the earth relative to the star. The result was negative, as Mascart later pointed out it should have been in any case, but it led Fresnel to consider the question of the passage of light through moving bodies. He formed the hypothesis that the æther inside a moving body is only carried with the body in so far as the latter affects its properties; and this led him to the conclusion that if u be the velocity of the body, the æther included in it moves with a velocity $u(\mu^2-1)/\mu^2$. This was one of several examples, where clear insight amounting to genius enabled Fresnel to grasp the essential features of a problem and obtain results that were experimentally correct though based on weak foundations. A celebrated experiment by Fizeau in 1851 confirmed Fresnel's formula. In this experiment a beam of light was sent through two parallel tubes in which water was kept flowing in opposite directions. The two parts of the beam were brought to interference and the change of refractive index caused by the motion could thus be measured. The experiment was confirmed by Michelson, but we give it now a different interpretation. The difference between a wave-velocity in a material body and that *in vacuo* is, as explained in Art. 150, not due to any alterations in the properties of the æther, but to an effect of the electric oscillations on the molecules of matter. Our equations assumed that matter to be stationary, and when appropriate changes are made in the equations so as to introduce an additional term due to its velocity, Lorentz has shown that Fresnel's results hold good. So far as Fizeau's experiments have any bearing on the subject they seem to confirm the hypothesis of a stationary æther.

196. The Michelson-Morley experiment. The experiment of Fizeau referred to in the previous article was limited in its application by the velocities which can be imparted to the flow of a liquid in tubes. The relation between the æther and bodies moving through it with planetary velocities involves questions of a different order. The experimental difficulties are increased accordingly, but Michelson conceived the idea that it would be possible to obtain, by means of his interferometer, some indication of the effects of the earth's velocity in space. After a preliminary series of measurements Morley joined him in the research, which has led to results of far-reaching importance.

Let the velocity u of an observer be in the direction SQ (Fig. 183), along which we also suppose a beam of light to be propagated. The velocity of light being c, its relative velocity to the observer is $c \mp u$ according as he moves with or against the light. If after reaching Q the light is reflected back to S, the total time occupied in the forward and return journey is:

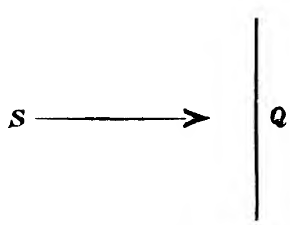


Fig. 183.

$$T_1 = L\left(\frac{1}{c+u} + \frac{1}{(c-u)}\right) = \frac{2Lc}{c^2 - u^2}.$$

If u be small compared with c, we may, neglecting powers of u/c higher than the third, write more conveniently:

$$T_1 = 2Lc^{-1}(1 + (u/c)^2).$$

Owing to the velocity u the optical length has therefore to be increased by $2Lu^2/c^2$.

If the motion be at right angles to SQ the time occupied in the journey cannot be determined in quite so simple a manner. In Michelson's experiment, a parallel beam of light falls on a mirror M_1M_2 (Fig. 184) inclined at an angle of 45° and we must trace the path of the reflected light if the whole apparatus moves in the direction of the incident light. As each of the wave-fronts such as AB moves forward it will intersect the mirror at successive points between M_2 and M_1 (Fig. 184), and by the time the wave would—if the mirror were at rest—

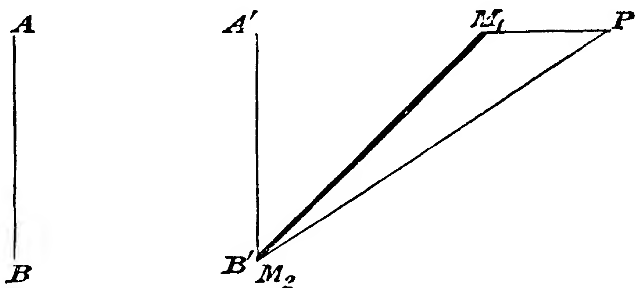


Fig. 184.

have reached M_1 that point has moved forward to some point P, and the wave is therefore reflected as if the mirror occupied the position M_2P. There is here an apparent violation of the law of reflexion, the reflected beam being turned through an additional small angle M_1M_2P which we proceed to calculate. Let A' be a point such that $A'M_2$ is parallel, and $A'M_1$ at right angles to the incident wave-front. Put $A'M_1 = a$, $M_1P = e$, and $A'M_2 = b$, while θ stands for the angle $A'M_2P$. We have to find the increment of the angle θ when a increases to $a+e$, the ratio of e to a being equal to $u/(c-u)$. If ϵ be the required angular increment we have $a = b\tan\theta$ and hence:

$$e = \delta a = \delta\theta\,(\tan\theta) = b\epsilon/\cos^2\theta$$

or

$$\epsilon = eb^{-1}\cos^2\theta = u\sin\theta\cos\theta/(c-u).$$

The beam of light being turned through twice the angle through which the mirror is rotated, it follows that when the angle of incidence is 45° the angle between the incident and reflected rays is increased by a quantity which to the first order of magnitude is equal to u/c. Let us return to Michelson's arrangement of mirrors, where S is the luminous point sending parallel rays in the direction Sa, and this is the direction in which we imagine the earth to be moving. The waves fall on the mirror a inclined at an angle of 45° (Fig. 185) and we must trace the rays as they travel referred to a system of coordinates *fixed in space*. The reflected ray ab is now turned through an angle greater by u/c than a right angle. As the mirror b travels parallel to itself, it reflects the light as if it were stationary and on its return the

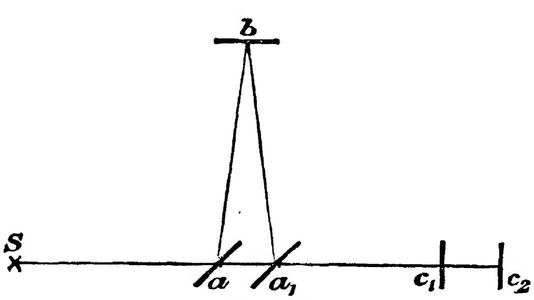


Fig. 185.

ray will intersect the line Sc at some point a_1. The total distance travelled, aba_1, is easily found by geometry to be

$$2L\,(1 + u^2/2c^2).$$

The time taken by the light during the passage is $c^{-1}L$. The angle between ab and a_1b being $2u/c$, the distance aa_1 is $2c^{-1}uL$. In consequence of its velocity u, the mirror therefore travels from a to a_1 in a time equal to that which it takes the reflected light to travel along the path aba_1. The ray reflected by the mirror c returning to a_1 will be reflected in the direction ba, and therefore overlap the ray returning from b.

We have now to determine the effect of the earth's motion, supposed to be in the direction Sc on the optical lengths which determines the position of the interference rings. We have found above that there is an increase in optical length in the forward and return path between these mirrors equal to $2Lu^2/c^2$ and an increase of half that amount in the path from a to b and back to the same mirror. The difference in the optical lengths which is effective in altering the position of the interference rings is therefore Lu^2/c^2, the light reflected from c having the longer path. For the purpose of the experiment the apparatus is mounted so that it can turn round a vertical axis, and if it be so turned through a right angle it will be the light reflected from b that now has the longer path. If the æther be stationary there should therefore be a shift of the interference bands corresponding to a difference of optical length amounting to $2Lu^2/c^2$, when the apparatus is rotated through a right angle.

No account is taken in the above reasoning of the velocity of the solar system in stellar space. Any effect of such a velocity would shew itself if the experiments were repeated at different seasons of the year.

We must next enquire into the order of magnitude of the expected effect. The direction of the earth's motion round the sun is towards a point of the ecliptic and at an angle of 90° westward of the sun, if we assume for the sake of simplicity that the earth's orbit is a circle. It will therefore have its full value u when that point is on the horizon, which takes place about noon, varying to some extent with the seasons of the year. As the ratio u^2/c^2 is about 10^{-8} it follows that when the expected difference of optical paths is equal to the tenth of a wavelength of sodium light $2L \times 10^{-8} = 6 \times 10^{-6}$ or that the distance L must be not less than 3 metres. Michelson carried out some experiments in 1881, the distance L being 1·2 metres. No displacement could be observed, but the result was not decisive as the expected shift only amounted to ·04 of the distance between two fringes. The experiments were repeated with improved apparatus by Michelson and Morley*.

* *Phil. Mag.* Vol. xxiv. p. 449 (1887).

To secure steadiness during the rotation the apparatus was mounted on a massive stone floating in mercury and the length of the path L was increased to 11 metres with the help of suitably disposed auxiliary mirrors, which reflected the light backward. The expected displacement was 0·4 of a fringe, and the twentieth part of this could not have escaped detection. Nevertheless no displacement depending on the orientation of the apparatus could be observed. There seems at first sight no escape from the conclusion that the æther at the surface of the earth partakes in the planetary velocity of the earth, a conclusion which however leads to great difficulties in other directions. It became necessary therefore to examine whether the apparently obvious interpretation of the experiment did not contain some hidden assumption which destroyed its validity as a crucial test. On examining critically the process of the experiment we find indeed such an assumption which at the time seemed not only justified but axiomatic. The measurement consists in comparing the optical length of two distances one of which lies in the direction of the earth's planetary velocity and the other at right angles to it. The result depends on the shifting of the interference rings when the apparatus is turned through a right angle. The assumption tacitly made is, that the relative lengths are unaltered during the change in the orientation of the apparatus. In other words it is assumed that a measuring rod has the same length whether it lies in the direction of the earth's velocity or at right angles to it. The discussion of this question which affects our fundamental conceptions of space and time is outside the limits of this treatise. As a matter of history it has led to what is now known as the Principle of Relativity.

CHAPTER XV.

EMISSION SPECTRA AND THE QUANTUM THEORY.

197. Atomic angular momentum. The theoretical interpretation of emission spectra, and the reduction of their phenomena to a systematic order, have progressed very rapidly in recent years, since the application of the principles of the Quantum theory to the types of model atoms to which we have been led by the study of other branches of Physics. We shall, in the succeeding chapters, discuss only those aspects of the Quantum theory which are most germane to the study of Optics. For further information than can be given in the brief sketch which we contemplate the reader may be referred to such works as those of Jeans*, Sommerfeld†, and Bohr‡.

It is convenient, at the outset, to give a short summary of Bohr's original theory of the Hydrogen spectrum. Nicholson§, in a study of the lines of unknown origin in the spectrum of the solar corona, found it possible to ascribe them to a set of states of atoms of simple electrical constitution, in which the states were fixed by the fact that the angular momentum in an atom could only have discrete values. He pointed out that this was, in effect, an introduction of Planck's quantum of action h, into the mechanics of the atom, in the only manner in which it appears naturally to be relevant. For "action" is a product of energy and time, with the physical dimensions ML^2/T which are also the dimensions of angular momentum. These states, now known in Bohr's theory as "Stationary states," in fact appeared to be determined by the quantum h, for their angular momenta were simple multiples of it, in linear order.

The spectra mentioned are not, as is at present believed, emission spectra in the ordinary sense, capable of arrangement in the types of series we have already reviewed, but are of the nature of "resonance spectra." The fundamental line given by any such state has a wave length $2\pi C/\omega$, where C is the velocity of light, and ω is the angular

* *Report on Radiation and the Quantum Theory*, Phys. Soc. Lond. 1914.

† *Ann. der Physik*, LI. pp. 1—94 (1916), pp. 125—167 (1916); *Atombau und Spektrallinien*, Leipzig (1919).

‡ *Phil. Mag.* XXVI. pp. 1—25 (1913), pp. 476—502 (1913), and other works.

§ *Monthly Notices of R.A.S.* 1910 *et seq.*

velocity of the electrons, moving in a path presumed to be circular. We can at once find the curious law which such spectra follow. For if n electrons, $E_1 \dots E_n$, equally spaced and each of charge $-e$ and mass m, rotate in a circle of radius r about a fixed nucleus $+\nu e$ of positive electricity, placed at C, Fig. 186, the dynamical equation for any one of them is easily found to be

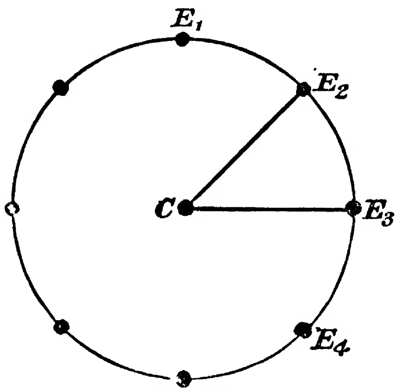


Fig. 186.

$$ma\omega^2 = \frac{e^2}{a^2}(\nu - \tfrac{1}{4}S_n),$$

where $$S_n = \sum_{r=1}^{n-1} \operatorname{cosec} \frac{r\pi}{n}.$$

Here m is the mass of an electron, and the total force on the electron has been resolved along the radius vector.

No other dynamical equation applies to the orbit, so that a, the radius, and ω, the angular velocity, can be found separately. This fact constitutes the difficulty emphasised by Lord Rayleigh, in regard to the construction of a theory of spectra. Until the advent of the Quantum theory of Planck, with its introduction of a new universal constant of nature having necessarily some atomic significance in its ultimate interpretation, there was nothing to indicate where to look for a second relation between a and ω. Nicholson's view, that h was essentially an angular momentum, allows us to write $ma^2\omega$ as some multiple of h, say

$$ma^2\omega = \tau h/2\pi.$$

Then with the preceding equation,

$$\omega = \frac{8\pi^3 me^4(\nu - \frac{1}{4}S_n)^2}{h^3\tau^3}$$

and $$\lambda = \frac{2\pi C}{\omega} = \frac{\tau^3}{2(\nu - \frac{1}{4}S_n)^2} \bigg/ \frac{2\pi^2 me^4}{h^3 C}.$$

With the values

$$e = 4{\cdot}774 \,.\, 10^{-10}, \quad h = 6{\cdot}545 \,.\, 10^{-27}, \quad e/m = 1{\cdot}767 \,.\, 10^7,$$

which involve the best recent determinations by Millikan, we find

$$\tau^3 = 2\lambda(\nu - \tfrac{1}{4}S_n)^2 \times 1{\cdot}097 \,.\, 10^5,$$

where $1{\cdot}097$ is in fact a simple sub-multiple of the Rydberg constant

$$N = 109679$$

of all series spectra.

The effect of this view is to give a simple set of spectrum lines in which $\lambda \propto \tau^3$, corresponding to the stationary states of any specific atomic structure with a definite nucleus νe and a definite number n of electrons. In other words, series should exist in which, as τ increases by equal steps, $\lambda^{\frac{1}{3}}$ forms a set of terms in an arithmetical progression. Such series have not been found in terrestrial elements, but perhaps because the conditions suitable for their production are not realized readily. But the spectrum of the corona contains many such arrangements involving practically all the lines (about 35 in number). We give one interesting example. The lines of wave-lengths 5303·3, 4359, 3534·5 Å.U., appeared to form such a series, for their cube roots are

$$\begin{array}{lll} \lambda = 5303{\cdot}3, & \lambda^{\frac{1}{3}} = 17{\cdot}439 & \\ & & 1{\cdot}104 \\ \lambda = 4359, & \lambda^{\frac{1}{3}} = 16{\cdot}335 & \\ & & 1{\cdot}102 \\ \lambda = 3534{\cdot}5, & \lambda^{\frac{1}{3}} = 15{\cdot}233 & \end{array}$$

The mean difference in cube root is 0·1103, and the next member towards the red is

$$\lambda^{\frac{1}{3}} = 18{\cdot}542, \quad \lambda = 6374{\cdot}8.$$

Some time after the discovery of these relations, a strong line at λ 6374·6 was discovered by Deslandres and Carrasco, in an eclipse expedition—a striking verification in a region of the coronal spectrum otherwise practically empty.

As to the origin of these lines, we have stated already that they probably belong to resonance spectra, being caused by the small vibrations—some of which the theory of Bohr allows—of a stationary state. They can only become perceptible on a photographic plate if the states persist as such for a long period, while the atoms are actually excited and not ionized, or if the light under examination comes from an enormous thickness of gaseous material. The realization of these conditions in the laboratory would be very difficult. The energy emitted in these radiations is not, as in the theory of emission spectra, the intra-atomic energy of the atom, but is derived from external sources, such as the ordinary solar radiation into the corona, taken up and emitted with these periodicities while the atom is seeking to return to its stationary state. In this sense the word resonance is used. The atoms are not in the usual condition which the quantum theory developed later appears to indicate, in that the orbits there are rarely circular. But these resonance frequencies may yet persist with ω as a mean angular velocity.

198. Bohr's theory and the spectrum of Hydrogen. We proceed to describe Bohr's theory as it was originally presented for the spectra of Hydrogen and of charged Helium. The theory is designed to account for a real emission spectrum, in which the energy of the emitted waves is definitely atomic before its emission, and a necessary adjunct of the stationary state immediately preceding the emission.

Bohr's first paper was published* shortly after the work described in the preceding section, and it adopted the hypothesis of a set of configurations for a given atom, and their specification in terms of angular momentum of electrons, but laid down the definite criterion that the angular momentum of every electron in every atom is $\tau h/2\pi$, where τ is an integer—a suggestion now known to have restricted validity. The orbits in the simple problem of Hydrogen—regarded as a single electron and a single nucleus of charge $+e$—and of positively charged Helium —regarded as an electron and a nucleus of charge $+2e$—were taken as circular.

If we consider an electron and a nucleus νe, of masses m, M respectively, the former being about 1/1835 of the latter, their orbital motion consists of rotations about the common centre of gravity O, Fig. 187, and the line joining them passes through O continually while each describes a circle.

Fig. 187.

We have $AO = \dfrac{m}{M+m} . AB$, $BO = \dfrac{M}{m+M} . AB$.

The equation of circular motion for the electron is

$$m . OB . \omega^2 = \frac{\nu e^2}{AB^2},$$

and for the nucleus $$M . OA . \omega^2 = \frac{\nu e^2}{AB^2}.$$

These are identical, and if $AB = r$,

$$mr^3 \omega^2 = \frac{\nu e^2}{M}(m + M) \quad \text{......................(1).}$$

The angular momentum of the system is

$$(m . OB^2 + M . OA^2)\, \omega.$$

If the angular momentum is always a multiple of a definite quantum we may write $\tau h/2\pi$ for it, where τ is an integer and h a constant. By reduction:

$$mr^2 \omega = \frac{\tau h}{2\pi M}(M + m) \quad \text{....................(2).}$$

Combining (1) and (2) we find

$$\omega = \frac{8\pi^3 m e^4 \nu^2}{h^3 \tau^3} \cdot \frac{M}{M+m},$$

$$r = \frac{h^2 \tau^2}{4\pi^2 m e^2 \nu} \cdot \frac{M+m}{M},$$

and $$\omega r = \frac{2\pi \nu e^2}{\tau h}.$$

* *loc. cit.*

The kinetic energy is

$$\tfrac{1}{2}(m\,.\,OB^2 + M\,.\,OA^2)\,\omega^2 = \tfrac{1}{2}mM(m+M)^{-1}\omega^2 r^2,$$

which takes the value $\dfrac{2\pi^2 me^4\nu^2}{h^2\tau^2}\cdot\dfrac{M}{M+m}$.

It is a general property of systems of electrical charges in steady motion under the inverse square law, that the potential energy is less than its value in a state of infinite dispersion by an amount equal to twice the kinetic energy. It follows that the energy which must be given to an atom to disperse it completely into its electrons and nucleus is equal to its kinetic energy. In the present instance, the potential energy is

$$C - \frac{\nu e^2}{r} = C - \frac{4\pi^2 me^4\nu^2}{h^2\tau^2}\cdot\frac{M}{m+M},$$

or $C - 2W$, where C is the potential energy in infinite dispersion, and W is the kinetic energy. Thus the variable part of the total energy is W, where

$$W = \frac{2\pi^2 me^4}{h^2\tau^2}\cdot\frac{M}{M+m}\cdot\nu^2,$$

and the energy lost in passage between two states τ_1 and τ_2, where τ_2 represents the final state, and $\tau_2 < \tau_1$, is

$$W_1 - W_2 = \frac{2\pi^2 me^4}{h^2}\cdot\frac{M}{M+m}\left(\frac{1}{\tau_1^2} - \frac{1}{\tau_2^2}\right)\nu^2.$$

Bohr regards this energy as being emitted in the form of a light pulse, connected with the frequency f by the quantum relation

$$W_1 - W_2 = hf.$$

In other words, the light pulse constitutes *one* quantum of emission. The frequency is therefore

$$f = \frac{2\pi^2 me^4\nu^2}{h^3}\cdot\frac{M}{M+m}\cdot\left(\frac{1}{\tau_1^2} - \frac{1}{\tau_2^2}\right).$$

The corresponding wave-length shown in the spectrum is $\lambda = \dfrac{C}{f}$, and the "wave-number" n, or number of complete waves contained in a length of one centimetre, is

$$n = \frac{10^8}{\lambda} = 10^8\,.\,\frac{2\pi^2 me^4\nu^2}{Ch^3}\;\frac{M}{M+m}\left(\frac{1}{\tau_1^2} - \frac{1}{\tau_2^2}\right).$$

Since τ_1 and τ_2 may take any integer values, it being presumed that the electron may change over from any stationary orbit to any other, this formula is already in qualitative agreement with the Balmer formula for Hydrogen (Art. 197), which is given exactly by

$$n = 109678{\cdot}3\left(\frac{1}{\tau_1^2} - \frac{1}{\tau_2^2}\right).$$

But the agreement is in fact quantitative. If we neglect m/M, known to be of order 10^{-3}, and write $\nu = 1$, the Rydberg constant, whose actual value is 109678·3, should be given by

$$N = 10^8 . \frac{2\pi^2 me^4}{Ch^3} .$$

This is found by calculation to be correct, with the accepted values of e, m, h, to one part in 1000, and therefore within the limit of experimental error for these magnitudes. The success of the theory is therefore very striking at the outset. The main Hydrogen series is the Balmer series, with $\tau_1 = 2$, $\tau_2 = 3, 4, 5, \ldots$. This represents the falling in of the electron from various orbits to its "second" orbit characterized by angular momentum h/π.

If the electron falls into its "third" orbit, $\tau_1 = 3$, and the series

$$n = N\left(\frac{1}{3^2} - \frac{1}{\tau_2^{\,2}}\right)$$

is developed. This is known usually as the Ritz series, and lies in the infra-red. The third known series is that of Lyman, in the ultraviolet, practically at the limit in which observations are possible by existing methods. This corresponds to falls of the electron into its "normal" or first orbit $\tau_1 = 1$.

For a fall to $\tau_1 = 4$, we obtain a series yet further into the infra-red, which is beyond the possibility of observation at present.

The entire visible spectrum of the Hydrogen atom, when neutral electrically, is, by the present theory, contained in these sets of series, and all other lines of Hydrogen,—*i.e.* the whole "secondary" or "compound-line" spectrum, must be ascribed to other systems, such as the charged atom or the neutral or charged molecule.

199. The spectrum of ionized Helium. Helium is now regarded as the next element after Hydrogen, in order of atomic simplicity. Its ordinary state involves an atomic structure consisting of a nucleus $+2e$ of positive electricity—if complex, yet so confined as to act effectively like a point charge—and two electrons describing orbits round it. In its ionized condition, it has completely lost one electron, and the other was regarded by Bohr as moving in one of a set of possible circular orbits, as defined in the preceding section.

For its spectrum, we merely write $\nu = 2$, and the resulting wave-number should be

$$n = N' \left\{\frac{1}{(\tau_1/2)^2} - \frac{1}{(\tau_2/2)^2}\right\},$$

where N' is again Rydberg's constant, though modified slightly.

For if N is the value for Hydrogen, we clearly have

$$\frac{N'}{N} = \frac{M_1 + m}{M_1} \cdot \frac{M_2}{M_2 + m},$$

where M_1, M_2 are the masses respectively of a Hydrogen and of a Helium nucleus—effectively the masses of the atoms, so that $M_2/M_1 = 4$, the atomic weight of Helium.

By this formula, the theory can interpret some spectrum lines which had previously led to much controversy. Originally these lines were discovered by Pickering in the spectrum of the star ζ Puppis, and subsequently in other stars and the nebulæ. They were believed to be due to Hydrogen, on account of the simple relations between the formulæ, and were classified by Rydberg as the "sharp" and principal series of Hydrogen. The most important line was given by

$$n = N\left(\frac{1}{(1{\cdot}5)^2} - \frac{1}{\tau_2^2}\right)$$

with $\tau_2 = 2$, and the calculated wave-length 4688 Å. was believed to agree with observation. Its wave-length is now known accurately as 4685·81 Å. The second member of the series, $\tau_2 = 3$, has wave-length 2734 Å. below the limit of observation of stellar spectra, whose radiations are absorbed in this region by the terrestrial atmosphere. The other known lines were the "Pickering series," believed to follow the form

$$n = N\left(\frac{1}{2^2} - \frac{1}{(\tau_2 + 0{\cdot}5)^2}\right)$$

with wave-lengths 5412, 4541, 4200, ...

Clearly both these series are included, at least very closely, in the general formula we have developed for ionized Helium, but we notice that this general formula indicates the existence between any two consecutive lines of these latter series, of another member of the set, those which occur between the members of the last set being approximately coincident with the Balmer lines of Hydrogen. These lines, still believed to be due to Hydrogen, were ultimately obtained with sharp definition by Fowler, from a tube containing Helium and only a trace of Hydrogen. The line λ 4686 had previously been noticed by Lockyer and others, but never really identified as regards its origin. The conditions of production of the spectrum are now well known.

Fowler found a new strong line at λ 3203, and both this and λ 4686 appeared to form the first members of series. Fowler showed subsequently that the actual series formula which they really fitted is

$$n = 4 \times 109723{\cdot}22\left(\frac{1}{3^2} - \frac{1}{\tau_2^2}\right)$$

with the observed members, in International Ångström units in air:

τ_2	λ	τ_2	λ
4	4685·81	8	2385·46
5	3203·16	9	2306·18
6	2733·34	10	2252·72
7	2511·25		

The degree of fit is effectively absolute, and the lines cannot be regarded as forming two interlacing series, even from the point of view of their gradation of experimental intensity. They can now be obtained in pure Helium.

The corresponding Pickering series was also shown by Fowler to be

$$n = 4 \times 109723{\cdot}22 \left(\frac{1}{4^2} - \frac{1}{m^2}\right),$$

and seventeen members have been observed. The first six are:

τ_2	λ	τ_2	λ
5	10123·72	8	4859·36
6	6560·16	9	4541·63
7	5411·57	10	4338·71

The alternate members $\tau_2 = 6, 8, 10, \ldots$ lie between the lines observed by Pickering, and coincide more and more closely with the Balmer lines of Hydrogen, $\lambda\lambda$ 6562·79, 4861·33, 4340·47, ... in International Ångström units. They cannot be derived from Hydrogen, and only admit, with the whole set, the formula quoted.

The theory predicted also the appearance of an ultraviolet series

$$n = 4 \times 109723{\cdot}22 \left(\frac{1}{2^2} - \frac{1}{\tau_2{}^2}\right)$$

of which four members, $\lambda\lambda$ 1640·49, 1215·18, 1084·98, 921·39, corresponding to $\tau_2 = 3, 4, 5, 6$, have been found by Lyman.

200. The Rydberg constant. It appears from the previous article that N is not an absolute constant but slightly different for different spectra. Its values for Hydrogen and Helium, for example, have been shown to be related by

$$\frac{N_{\mathrm{He}}}{N_{\mathrm{H}}} = \frac{1 + \dfrac{m}{M_1}}{1 + \dfrac{m}{M_2}},$$

where M_1, M_2 are the masses of the Hydrogen and Helium atom respectively.

The accepted value for M_1/m was 1840 before the advent of Bohr's theory, and since, by atomic weight determinations, $M_2/M_1 = 4/1{\cdot}008$, the value of N_{He} can be calculated from N_{H}, which was well known. The result is in precise agreement with the value deduced by Fowler from his observations on the Pickering lines.

Fowler applied the argument in the converse direction, in his Bakerian lecture, using the values of N for Hydrogen and Helium to obtain the best estimate of the mass of the electron in relation to that of a Hydrogen atom. Paschen* has made a very precise estimate in this way, taking account of relativity corrections, and other small effects, with the result that

$$M_1/m = 1843{\cdot}7.$$

The true Rydberg constant, which is a multiplier of the wave-numbers of all spectra, may be called N_∞, being the value belonging to a nucleus of infinite mass. Its value is found to be

$$N_\infty = 109737{\cdot}7,$$

and this is practically the value for all but very light elements.

201. Enhanced Series. The lines of ionized Helium are an example of what Fowler has called a "$4N$" series, or, in the previous terminology, a spark or enhanced series. Such spectra are emitted under spark conditions in a vacuum tube—though with various degrees of ease according to the element under examination, and if emitted under the simple arc discharge, they become relatively brighter in the spark, in which the arc lines become weaker. While, in a vague way, it was beginning to be realized that they probably arose from atoms which had permanently lost one or more electrons, it remained for Fowler, as a deduction from Bohr's theory, to lay down the precise quantitative significance of such series in the general scheme of spectra.

When one electron has been driven from an atom, the rest of the atom, being endowed with charge $+e$, behaves at a sufficient distance like a Hydrogen nucleus. The lost electron, in its return, is approximately under the influence of such a nucleus, and emits, in passing to an inner state, a line for which the Rydberg constant is a multiple of N—being, apart from correction of order m/M, $N\nu^2$ for a nucleus of charge νe.

This is the ordinary, or arc type of emission, and corresponds to the simplest operation of the discharge in removing one electron, or simple ionization. For double ionization, the more energetic spark conditions are needed, two electrons are driven off, and the return of one of them to an atom behaving approximately like a Helium nucleus gives series spectra with 2^2N as the Rydberg constant. Other examples are known in addition to Helium, and more especially the alkaline earths.

* *Ann. der Physik*, L. p. 935 (1916).

The spark spectrum of magnesium, for instance, has been arranged in an elegant set of such series by Fowler.

With yet higher degrees of excitation, "$9N$" or 3^2N series are to be expected, though none as yet have been definitely isolated, from the few cases in which "super-spark" spectra are at present known.

202. The Combination Principle. A fundamental advantage of the quantum theory of spectra is the account it gives of the combination principle, already described. The value of W in more complex atoms will necessarily be a function of a quantum integer in any state, and the frequency of a line is always the difference of two such functions. By the localization of these functions, one to each possible state of the atom, the theory has achieved a somewhat convincing account of the actual relations between spectra, which has been very remote from all previous methods of regarding the problem. But it has not, as yet, given an indication of the *physical* relationship between the principal and subordinate series of any given element.

203. Roentgen Rays. The study of Roentgen radiation has not yet revealed, with certainty, elaborate series relations such as are found in visible spectra, but there are many indications of their existence. The most prominent "characteristic" radiation for any atom—a natural type of radiation emitted from the atom concerned—consists of what are known as the K-lines and the L-lines. Current atomic theory relates such radiations to rings of electrons very close to the nucleus of an atom in comparison with the electrons producing series spectra. This theory allows a certain amount of mathematical treatment.

Moseley* first proved that the wave-numbers of the fundamental, or strongest, line from each element were proportional to the squares of a quantity which increased by unit steps in passing from one atom to the next in the periodic table. The magnitude involved is clearly the atomic number of the element, or the charge on its nucleus. The wave-numbers for the K-lines are proportional to $(N-1)^2$, and those of the L-lines to $(N-7 \cdot 4)^2$. There is thus a clear correspondence with the spectrum, given by the principles of the quantum theory, of a ring of electrons. For a ring of n electrons surrounding a nucleus ve, the spectrum would be, N here denoting the Rydberg constant,

$$n = N\left(\nu - \tfrac{1}{4}S_n\right)^2 \left\{\frac{1}{\tau_1^2} - \frac{1}{\tau_2^2}\right\},$$

where S_n is usually incommensurable. The atomic number of the atom would be ν. From the fact that S_n appears to be constant, it would seem that there is a fairly permanent type of constitution in the most internal part of the atom.

* *Phil. Mag.* LX. April, 1914.

Further important evidence from Roentgen spectra is contained in an investigation by Sommerfeld. As we have seen already, the effect of the change of mass of an electron with speed is to produce a structure in ordinary visible spectrum lines. The main components in each line of the Balmer series are two in number, and the effect produced is a doublet. If Δ_H is its separation, in wave-number, an investigation for an electron moving round a nucleus νe instead of e shows that the general separation is $\nu^4 \Delta_H$. This rapidly increases with ν and we ultimately, for heavier elements, obtain enormous separations, so that the doublets overlap and cease to be recognized as such except by their numerical relations.

Such considerations should apply also to the Roentgen spectra, and the K and L series should be doublets whose separations are simply related to those in the hydrogen spectrum.

In fact, if Δ is the observed separation in the K or L series of any element of atomic number ν, Δ/ν^4 should be constant. This relation is confirmed very strongly, as for example, in the sequence of elements

Chromium, Copper, Zinc, Bromine, Rhodium, Silver

with well measured K spectra, where the values are

0·43, 0·40, 0·39, 0·47, 0·35, 0·44.

As the elements get heavier, the formula must be regarded as an approximation, and small discrepancies, gradually increasing, are to be expected.

204. The radiation of a black body. In the preceding articles it has been shown how by means of the fundamental ideas of the quantum theory and certain plausible *ad hoc* assumptions, we could arrive at an equation which represented the characteristic disposition of spectroscopic lines in a series spectrum, and in the case of hydrogen even led quantitatively to the correct wave-lengths. The success of the theory is enhanced by the fact that it was originally introduced by Planck with an entirely different object in view, that object being the correct representation of what is called the Black Body Radiation. The Principle of the Conservation of Energy leads to the conclusion that inside an enclosure of uniform temperature there is a perfectly definite distribution of intensity between the different frequencies depending only on the temperature, which is identical with the distribution of intensities in the radiation of a perfectly black body. If the laws of classical dynamics be applied to molecular thermal motion, it was shown by Boltzmann as an extension of a theorem proved by Maxwell that the total energy is equally divided between all degrees of freedom, and proportional to the absolute temperature. It is clear that with regard to the molecular velocities which determine radiation this equipartition of energy cannot hold. In the first place, the number of degrees of freedom for frequencies which range from zero to infinity is infinite, and therefore the total radiation would be infinite, and in the second place the total radiation is found to be proportional to the fourth instead of the first power of absolute temperature. It is remarkable, however, that the law of equipartition leads—as shown by Jeans and Rayleigh—to a formula which is correct for waves of small frequencies. It is in order to find a theoretical basis for the experimental facts that Planck introduced the quantum theory according to which the total oscillatory energy consists of integer multiples of a quantum which is proportional to the frequency.

Planck has stated his theories in two different forms, both of which present difficulties. The second is now more generally used, but it contains the implication that while the emission of radiation by an atom is discontinuous, the absorption is continuous. Wilson* has given a treatment of the black body formula which avoids this apparent defect, and which reconciles the Planck formula for complete radiation with the precise specifications of general quantum theory as used in the discussion of emission spectra. As this appears to be the most satisfactory account yet published, we shall follow it in detail. The analysis is closely parallel to the original analysis of Planck.

We begin with the hypothesis that all interchanges of energy between dynamical systems and the æther are discontinuous. The systems are conservative in general, but subject to short intervals in which definite energies can be absorbed or radiated. While they are conservative systems, they are ruled by ordinary dynamical laws and are said to be in steady states.

If $(q_1, q_2, \ldots)$ are the coordinates which specify the systems, and $(p_1, p_2, \ldots)$ are the corresponding momenta, let T be the kinetic energy expressed in terms of the q's and $\dot{q}$'s. If the q's are "normal" coordinates

$$T = \tfrac{1}{2}A_1\dot{q}_1^2 + \ldots + \tfrac{1}{2}A_n\dot{q}_n^2,$$

where the A's are explicit functions of the q's.

Since T is a homogeneous quadratic in the velocities

$$\begin{aligned}2T &= \dot{q}_1\frac{\partial T}{\partial \dot{q}_1} + \ldots + \dot{q}_n\frac{\partial T}{\partial \dot{q}_n}\\ &= p_1\dot{q}_1 + p_2\dot{q}_2 + \ldots + p_n\dot{q}_n.\end{aligned}$$

The typical kinetic energy, for a variation of only one q, say q_r, is

$$\tfrac{1}{2}A_r\dot{q}_r^2 = T_r \text{ (say)}.$$

Then

$$T_1 = \tfrac{1}{2}A_1\dot{q}_1^2, \quad T_2 = \tfrac{1}{2}A_2\dot{q}_2^2, \quad \ldots.$$

* *Phil. Mag.* 1914 and 1915.

When we are dealing with a conditionally periodic system, in which the various coordinates repeat their values in periods $\frac{1}{\nu_1}, \frac{1}{\nu_2}, \ldots$, we have at once

$$2 \int T_1 dt = \int p_1 dq_1$$

and similar equations for each coordinate, for

$$p_1 = \frac{\partial T}{\partial \dot{q}_1} = A\dot{q}_1.$$

Making, now, the supposition that, when integrated over a complete period of the corresponding coordinate,

$$\int p_r dq_r = \rho_r h,$$

where ρ_r is an integer, and denoting these integrals by H_1, H_2, ... we have the quantum specification:

For any system, let the H's be $H_1 = \rho h$, $H_2 = \sigma h$, $H_3 = \tau h$, ... where $(\rho, \sigma, \tau, \ldots)$ are its characteristic quantum integers in the steady state in which it happens to be at any time.

We proceed to consider the statistical equilibrium of a set of N systems, characterized by various groupings of the integers $\rho, \sigma, \tau, \ldots$. Let $N_{\rho,\sigma,\tau,\ldots}$ be the number which have $\rho, \sigma, \tau, \ldots$ as their integers, and therefore have also $H_1 = \rho h$, $H_2 = \sigma h$, ... and so forth. The proportion which have $\rho, \sigma, \tau, \ldots$ as integers, is

$$f_{\rho,\sigma,\tau,\ldots} = \frac{N_{\rho,\sigma,\tau,\ldots}}{N},$$

and the sum of all these fractions, taken for all possible values of all integers, is unity. If $E_{\rho,\sigma,\tau,\ldots}$ is the energy of a system as a function of its integers, the total energy is E, where E is the sum of all possible values of $NE_{\rho,\sigma,\tau,\ldots} f_{\rho,\sigma,\tau,\ldots}$.

The number of ways, P, in which N systems can be distributed, so that $N_{\rho,\sigma,\tau,\ldots}$ are of type characterized by $\rho, \sigma, \tau, \ldots$ is

$$P = \frac{N!}{\varpi\left(N_{\rho,\sigma,\tau,\ldots}!\right)},$$

where the symbol ϖ indicates the product of all possible values of the expression in the bracket, *i.e.* of the factorials of all possible numbers N characterized by any set of integers.

When there is a large collection of systems, all the numbers N will be large, and we may use Stirling's formula

$$n! = e^{-n} n^n \sqrt{2n\pi}$$

for each, and find

$$P = \frac{N^N}{\varpi\left(N_{\rho,\sigma,\ldots}\right)}.$$

Planck calls the number of ways of distributing a set of systems in a given way the "thermodynamic probability" of the given way, and by thermodynamic principles, if P is this probability, the entropy of the way is

$$\phi = k \log P,$$

where k is the entropy constant.

The entropy is accordingly

$$\phi = k\{N \log N - \Sigma N_{\rho,\sigma,\tau,\ldots} \log N_{\rho,\sigma,\tau,\ldots}\},$$

the summation being for all values of ρ, σ, τ, ..., and therefore a multiple summation. This is identical with

$$\phi = -Nk\, \Sigma f_{\rho,\sigma,\tau,\ldots} \log f_{\rho,\sigma,\tau,\ldots},$$

with the same meaning of the summation.

Statistical equilibrium occurs when

$$\delta\phi = 0,$$

the variation being subject to the constancy of total energy, and also to the fact that

$$\Sigma f_{\rho,\sigma,\tau,\ldots} = 1.$$

If we apply the ordinary method of undetermined multipliers, we have at once

$$1 + \log f_{\rho,\sigma,\tau,\ldots} + \beta E_{\rho,\sigma,\tau,\ldots} + \gamma = 0,$$

where β and γ do not depend on ρ, σ, τ, ... Thus

$$f_{\rho,\sigma,\tau,\ldots} = Ae^{-\beta E_{\rho,\sigma,\tau,\ldots}},$$

where $$1 = \Sigma\, Ae^{-\beta E_{\rho,\sigma,\tau,\ldots}}.$$

By the ordinary thermodynamic principles, as in Planck's theory, $\beta = 1/kT$, where T is the absolute temperature. We outline the proof very briefly.

We have

$$\begin{aligned}\phi &= -kN \Sigma f_{\rho,\sigma,\tau,\ldots} \log f_{\rho,\sigma,\tau,\ldots}\\ &= -kN \log A + kN\beta\, \Sigma\, AE_{\rho,\sigma,\tau,\ldots}\, e^{-\beta E_{\rho,\sigma,\tau,\ldots}}\\ &= -kN \log A + k\beta E,\end{aligned}$$

where E is the total energy, or $N\Sigma\, E_{\rho,\sigma,\tau,\ldots} f_{\rho,\sigma,\tau,\ldots}$.

Thus $$\frac{d\phi}{d\beta} = -\frac{kN}{A}\frac{dA}{d\beta} + kE + k\beta\frac{dE}{d\beta}.$$

But since $$A\, \Sigma\, e^{-\beta E_{\rho,\sigma,\tau,\ldots}} = 1,$$

we have $$0 = \frac{1}{A}\frac{dA}{d\beta} - A\, \Sigma\, E_{\rho,\sigma,\tau,\ldots}\, e^{-\beta E_{\rho,\sigma,\tau,\ldots}} = \frac{1}{A}\frac{dA}{d\beta} - \frac{E}{N},$$

and $d\phi/d\beta$ becomes $$\frac{d\phi}{d\beta} = k\beta\frac{dE}{d\beta},$$

so that $$\frac{d\phi}{dE} = k\beta,$$

and $k\beta$ is the reciprocal of the absolute temperature. The statistical distribution of the systems is therefore represented by

$$f_{\rho, \sigma, \tau, \ldots} = e^{-\frac{E_{\rho, \sigma, \tau, \ldots}}{k\tau}} \Big/ \Sigma e^{-\frac{E_{\rho, \sigma, \tau, \ldots}}{k\tau}}.$$

In order to obtain a form of Planck's theory as a special case, and the only form which is consistent with the quantum principles designed for spectra, we consider an elementary oscillator of Planck's type, which, in a steady state, moves according to the equation

$$m\ddot{q} + kq = 0.$$

If its frequency is ν, the solution is

$$q = R \cos (2\pi\nu t - \theta),$$

where R and θ are constants. Thus

$$p = -2\pi m\nu R \sin (2\pi\nu t - \theta),$$

and its energy is

$$2\pi^2\nu^2 mR.$$

Over a complete period, the phase integral is

$$\int_t^{t+\frac{1}{\nu}} pdq = 2\pi^2\nu mR^2 = \rho h,$$

and if E_ρ is the energy when the quantum number is ρ, we have at once

$$E_\rho = \rho h\nu,$$

where ν is the frequency. The energy is a whole number of quanta. The law of distribution of the systems with this single variable integer ρ is, by the last section,

$$f_\rho = e^{-\frac{\rho h\nu}{kT}} \Big/ \sum_{\rho=0}^{\infty} e^{-\frac{\rho h\nu}{kT}}$$

$$= \left(1 - e^{-\frac{h\nu}{kT}}\right) e^{-\frac{\rho h\nu}{kT}}.$$

This is Planck's law of distribution.

The average energy of an oscillator in such an assemblage is then readily found to be

$$\bar{E} = \frac{h\nu}{e^{\frac{h\nu}{kT}} - 1}$$

which is also a fundamental formula of Planck.

The æther, on Planck's theory, is to be looked upon as a collection of oscillators which interchange energy with each other, through the intervention of matter. In a frequency range from ν to $\nu + \delta\nu$, various writers* have shown that the number of these per unit volume is

$$\frac{8\pi\nu^2}{c^3} d\nu,$$

* *E.g.* Jeans, *Phil. Mag.* x. p. 91 (1905).

where c is the velocity of light. Attaching to each one the average energy above, corresponding to maximum entropy or the most probable distribution, we find that the total energy within the range ν to $\nu + d\nu$ is given by

$$U_\nu \, d\nu = \frac{8\pi h\nu^3}{c^3} \frac{d\nu}{e^{\frac{h\nu}{kT}} - 1}.$$

This is Planck's law of complete radiation, now definitely established.

CHAPTER XVI.

DYNAMICAL THEORY OF SPECTRA.

205. Extension of Bohr's Theory. We proceed to apply and extend the results of the last Chapter to the investigation of the spectrum given by an electron moving in any possible manner about a stationary nucleus charged with ν units of positive electricity e, and to show that it still leads, in the case of Hydrogen, to the Balmer formula alone, and in the case of charged Helium, to the Pickering lines alone. We use ordinary dynamics and the phase-integral specification of the quantum conditions which replace the "initial conditions" of orbits in ordinary dynamics. Essentially the investigation is that of Sommerfeld, though modified in some of the details, and more especially extended to three dimensions.

If we use spherical polar coordinates, with their origin at the nucleus, the kinetic and potential energies are

$$T = \tfrac{1}{2}m\,\{\dot{r}^2 + r^2\dot{\theta}^2 + r^2\sin^2\theta\,\dot{\phi}^2\},$$

$$V = -\frac{\nu e^2}{r},$$

and the equation of energy is

$$\tfrac{1}{2}m\,\{\dot{r}^2 + r^2\dot{\theta}^2 + r^2\sin^2\theta\,\dot{\phi}^2\} - \frac{\nu e^2}{r} = -W,$$

where $-W$ is the whole negative energy. The momenta are

$$p_1 = \frac{\partial T}{\partial \dot{r}} = m\dot{r}, \qquad p_2 = \frac{\partial T}{\partial \dot{\theta}} = mr^2\dot{\theta}, \qquad p_3 = \frac{\partial T}{\partial \dot{\phi}} = mr^2\sin^2\theta\,\dot{\phi},$$

and thus, if we eliminate the velocities,

$$p_1^2 + \frac{p_2^2}{r^2} + \frac{p_3^2}{r^2\sin^2\theta} - \frac{2m\nu e^2}{r} = -2mW.$$

Moreover, p_3 is constant—it is the condition, valid on ordinary dynamics, of constancy of angular momentum—and its phase integral is

$$\int_0^{2\pi} p_3\,d\phi = n_3 h,$$

where n_3 is an integer. We therefore have

$$p_3 = n_3 h/2\pi.$$

The equation of energy is one to which Jacobi's general theorem of "separation of variables" is applicable, and the integrals of the problem are, if β is constant,

$$p_2^2 + \frac{p_3^2}{\sin^2\theta} = \beta^2,$$

$$p_1^2 + \frac{\beta^2}{r^2} = \frac{2m\nu e^2}{r} - 2mW.$$

Thus the phase integral for p_2 becomes

$$n_2 h = \int p_2 d\theta = \int \sqrt{\beta^2 - \frac{p_3^2}{\sin^2\theta}}\, d\theta.$$

In order to obtain a real result, θ must range from

$$\theta = \sin^{-1}\left(\frac{p_3}{\beta}\right) \text{ to } \theta = \pi - \sin^{-1}\left(\frac{p_3}{\beta}\right)$$

and back again, with, however, a negative sign to p_2 on the return journey. Thus

$$n_2 h = 2\int_{\psi}^{\pi-\psi} \sqrt{\beta^2 - \frac{p_3^2}{\sin^2\theta}}\, d\theta,$$

where $p_3 = \beta\sin\psi$. The substitution

$$\beta^2\sin^2\theta = \beta^2\cos^2\phi + p_3^2\sin^2\phi$$

reduces this to

$$n_2 h = 4\int_0^{\frac{\pi}{2}} \frac{(\beta^2 - p_3^2)^2\sin^2\phi\cos^2\phi\, d\phi}{\beta^2\cos^2\phi + p_3^2\sin^2\phi},$$

or, if $\tan\phi = t$,

$$n_2 h = 4(\beta^2 - p_3^2)^2 \int_0^{\infty} \frac{t^2\, dt}{(1+t^2)^2(\beta^2 + p_3^2 t^2)}.$$

This integral is of a well-known type, and is readily evaluated. The final result is

$$n_2 h = 2\pi(\beta - p_3).$$

But $2\pi p_3 = n_3 h$, and therefore

$$\beta = (n_2 + n_3)\frac{h}{2\pi}.$$

This is the first significant result,—the angular quantum numbers are merely additive, a fact which could not have been foreseen before the analysis.

Reverting to the equation for p_1, we have

$$p_1^2 = \frac{2m\nu e^2}{r} - \frac{\beta^2}{r^2} - 2mW$$

and the radial quantum condition is

$$n_1 h = \int p_1 dr = \int dr \sqrt{\left\{-2mW + \frac{2m\nu e^2}{r} - \frac{\beta^2}{r^2}\right\}},$$

between the extreme values of r and back again, with a negative sign on the return journey. The form of the quadratic shows that the square root is only real if r lies between two positive values p and q. In fact

$$n_1 h = 2\sqrt{2mW}\int_p^q \sqrt{\left(1-\frac{p}{r}\right)\left(\frac{q}{r}-1\right)}\, dr,$$

where
$$q - p = \frac{\nu e^2}{W}, \quad qp = \frac{\beta^2}{2mW}.$$

This integral is evaluated readily by the substitution

$$r = p\sin^2\theta + q\cos^2\theta,$$

and we obtain, by simple analysis,

$$n_1 h = \frac{2\pi me^2\nu}{\sqrt{2mW}} - 2\pi\beta.$$

Accordingly, with the preceding value of β,

$$W = \frac{2\pi^2 me^4\nu^2}{h^2} \cdot \left(\frac{1}{n_1+n_2+n_3}\right)^2,$$

so that, in the final formula for W, all the quantum integers are additive, and count as one integer so far as the positions of the spectrum lines are concerned. It is very remarkable that the radial quantum number should be additive with those of the transverse quanta.

If N is Rydberg's constant, the wave-numbers (n) of all the possible lines are given by

$$n = N\nu^2\left\{\frac{1}{(n_1+n_2+n_3)^2} - \frac{1}{(m_1+m_2+m_3)^2}\right\},$$

where the n's and m's can take all possible integer values. No line is included which is not given by Bohr's simple theory. This investigation constituted the first success in the substitution of a uniform generalization for Bohr's original suppositions, though only this brief account appears to be necessary here.

For paths in one plane, however—which are, of necessity, ellipses—we may give a very simple illustration of the effect of applying this principle. The nature of the restriction of the orbits is, in fact, primarily a restriction of *eccentricity*.

This can be shown quite simply without reference to the preceding analysis. For in polar coordinates, let the equation of the ellipse be

$$\frac{l}{r} = 1 + e\cos\theta.$$

Then r and θ are the defining variables, and

$$m\int\dot{r}\,dr = n_1 h, \quad m\int r^2\dot{\theta}\,d\theta = n_2 h.$$

Now
$$-\frac{l\dot{r}}{r^2} = -e\sin\theta\,\dot{\theta},$$

so that
$$m\dot{r} = e\sin\theta \,.\, mr^2\dot{\theta}.$$

Since $r^2\dot{\theta}$ is constant, we have

$$mr^2\dot{\theta} = \frac{n_2 h}{2\pi}, \quad n_1 h = \int m\dot{r}\,dr = \int e \sin\theta \,.\, dr \frac{n_2 h}{2\pi}$$

Thus
$$n_1 h = \frac{n_2 h e^2}{2\pi}\int_0^{2\pi} \frac{\sin^2\theta\, d\theta}{(1+e\cos\theta)^2}$$
$$= n_2 h \left\{\frac{1}{\sqrt{(1-e^2)}} - 1\right\}.$$

Accordingly
$$\sqrt{(1-e^2)} = n_2/(n_1+n_2),$$
specifying all the eccentricities which are possible.

The quantum theory takes the place of the more usual "initial conditions" which would determine this eccentricity. From many points of view, perhaps the best view of the nature of the quantum conditions is found in this replacement.

206. Fine structure of spectrum lines. According to the relativity theory, as well as the dynamics of the Lorentz electron, if the mass of the electron for slow speeds is m, its kinetic energy when moving with velocity v is

$$T = mc^2 \left\{\frac{1}{\sqrt{\left(1-\frac{v^2}{c^2}\right)}} - 1\right\},$$

where c is the velocity of light. This of course becomes $\frac{1}{2}mv^2$ when terms of relative order $\frac{v^2}{c^2}$ are ignored.

The accurate equation of energy in the preceding section should therefore be, since

$$v^2 = \dot{r}^2 + r^2\dot{\theta}^2 + r^2\sin^2\theta\,\dot{\phi}^2,$$

and since the momenta are, on relativity principles,

$$p_1 = \frac{m\dot{r}}{\sqrt{\left(1-\frac{v^2}{c^2}\right)}}, \quad \ldots,$$

of the form, on reduction

$$-mc^2 + mc^2\left\{1 + \frac{1}{mc^2}\left[p_1^2 + \frac{p_2^2}{r^2} + \frac{p_3^2}{r^2 s^2\theta}\right]\right\}^{\frac{1}{2}} - \frac{\nu e^2}{r} = -W.$$

This equation again admits separation of variables, after the manner of Jacobi. Thus we write

$$p_3 = \text{constant} = \frac{n_3 h}{2\pi},$$

$$p_2^2 + \frac{p_3^2}{\sin^2\theta} = \beta_2^2,$$

where β is constant. If we quantize p_2, since the equation last written is the same as before, we again find the same values of β, namely

$$\beta = (n_2 + n_3)\frac{h}{2\pi},$$

involving additive angular integers.

The equation for p_1 is then

$$-mc^2 + mc^2\left\{1 + \frac{p_1^{\,2} - \beta^2/r^2}{mc^2}\right\}^{\frac{1}{2}} - \frac{\nu e^2}{r} = -W$$

so that

$$p_1^{\,2} = -\frac{\beta^2}{r^2} + m^2c^2\left\{-1 + \left(1 + \frac{\dfrac{\nu e^2}{r} - W}{mc^2}\right)^2\right\}.$$

It is sufficient, in a consideration of fine structure in the cases of Hydrogen and charged Helium, to retain only the first order correction. We therefore write, on reduction, neglecting terms containing c^{-4} and higher powers,

$$p_1^{\,2} = -H + \frac{2K}{r} - \frac{L^2}{r^2},$$

where

$$H = 2m\,W\frac{W^2}{c^2}, \quad K = me^2\nu - \nu e^2\frac{W}{c^2}, \quad L^2 = \beta^2 - \frac{\nu^2 e^4}{c^2}$$

and the quantization of p_1 gives

$$n_1 h = 2\int dr\,\sqrt{\left\{-H + \frac{2K}{r} - \frac{L^2}{r^2}\right\}}$$

between the extreme values of r making the square root real. We find easily that

$$n_1 h = 2\pi\left\{\frac{K}{\sqrt{H}} - L\right\}.$$

Substituting the values of K, H, L, and expanding in inverse powers of c, to the order involved already,

$$\frac{n_1 h}{2\pi} = \frac{m\nu\, e^2}{\sqrt{(2mW)}}\left\{1 - \frac{3W}{4me^2}\right\} - \beta\left\{1 - \frac{\nu^2 e^4}{\beta^2 c^2}\right\}.$$

The first approximation, with $\beta = (n_2 + n_3)\dfrac{h}{2\pi}$, is naturally

$$W = \frac{2\pi^2 me^4\nu^2}{h^2}\cdot\frac{1}{(n_1 + n_2 + n_3)^2}$$

and this may be used in determining the second, which becomes

$$W = \frac{2\pi^2 me^4\nu^2}{h^2}\cdot\frac{1}{(n_1 + n_2 + n_3)^2}\left\{1 + \frac{4\pi^2 e^4}{h^2c^2}\cdot\frac{\nu^2}{(n_1 + n_2 + n_3)^2}\cdot\frac{n_1 + \frac{1}{4}(n_2 + n_3)}{n_2 + n_3}\right\}.$$

The fine structure of the lines follows from the large variety of choice of the integers n_1 and $n_2 + n_3$, the latter being purely additive and equivalent to a single integer. For this value of W is no longer symmetrical in the integers defining the radial and transverse quanta. The validity of the formula has been confirmed most especially by Paschen* when $\nu = 2$. He examined the fine structure of many of the Pickering lines and found a variety of components for each, whose separations were very closely in accord with those to be expected from this formula when all possible values were put for the integers in the formula

$$W - W' = hf,$$

where f is the frequency of the emitted line.

This was a remarkable success for the quantum theory. The structure of the Hydrogen lines is more difficult to determine experimentally, owing to the extremely small separations. For it will be noticed that the separations are of order

$$\frac{4\pi^2 e^4}{h^2 c^2} \cdot \nu^2$$

proportional to ν^2, and are therefore, in the case of Hydrogen, roughly only about a quarter of the corresponding values for charged Helium.

Sommerfeld's own investigation is much more complicated, and extensive, in certain respects. For example, in two dimensions, he discusses the path of the particle, which is an ellipse whose perihelion moves round in the relativistic manner, so that the motion is not periodic. Doubt would thus be raised as to the precise manner in which the relation

$$\int p\, dq = nh$$

should be applied to the angular coordinate, if we did not take the specification of its application to pseudo-periodic systems defined earlier. We have here a definite distinction between the investigations of Wilson and Sommerfeld. The latter states that the principle, as he uses it, is purely tentative also in this respect—the appropriate limits for θ in the integration—and attempts to justify it by reference to Planck's cell-quanta, but is clearly not satisfied.

In Wilson's specification, the limits are quite precise, and are zero and 2π for θ.

The equation to the path is of the form

$$\frac{l}{r} = 1 + e \cos \mu\phi,$$

where μ is nearly unity, but in general incommensurable. This can be regarded as an ellipse whose perihelion slowly advances, and Sommerfeld's procedure was to integrate over a complete period 2π, with the justification that the advance of perihelion is extremely small.

* *loc. cit.*

The results of the two methods of looking at the problem are the same within any possible limits of experimental error, though Wilson's method is more precise and satisfactory.

207. The constant of spectral separation. Another interesting consequence of the last investigation is the emergence of a new universal constant of spectra, a constant of line-structure. For we may write

$$\alpha^2 = \frac{4\pi^2 e^4}{h^2 c^2}$$

and, numerically, inserting the accepted values of e and h,

$$\alpha = 7 \times 10^{-3}.$$

The fine structure of the lines of the Pickering series is in good accord with the view that α is a universal constant.

It should be pointed out, however, that while the quantum theory appears to give, correctly, the *positions* of the components of a fine structure as found in the laboratory, their relative intensities can vary greatly with circumstances. On the question of intensity, the theory has not yet proceeded far. We shall, at a later stage, indicate the direction in which speculation is proceeding, in a tentative way, towards a mode of prediction of the intensities and states of polarization of spectral lines.

The intensities, as we have stated, depend very much on the mode of excitation of the spectrum. For example, Merton and Nicholson, by interferometer measurements, were able to measure the separation in the lines H_α and H_β, then only showing two components. The separation of H_α confirmed Buisson and Fabry's value 0·132 Å.U., while that of H_β was 0·030 Å.U., the exact value appropriate to the case in which the Balmer series is regarded as a Principal Series, with separation vanishing at the limit of the series. This phenomenon occurred in the Hydrogen Spectrum under ordinary conditions. Merton has since shown that, under different conditions, a new and strong component appears. This may have been present weakly before, and there is no evidence against Sommerfeld's calculations of the *positions* of the components.

208. Dependence of fine structure on the mass formula. Other electrons than that of Lorentz have different formulæ for the variation of mass with speed, and it is interesting to know, although they violate the Principle of Relativity, the fine structure to which they lead in the spectrum. It is sufficient to mention an investigation by Glitscher*, who examined the Abraham electron from this point of view.

* *Ann. der Physik*, LII. pp. 608—630 (1917).

It was found that this electron, with its proper mass-formula, gave separations more than 10 per cent. too large. Glitscher regarded this fact, with Paschen's spectral measures, as a confirmation of the validity of the Lorentz electron. It seems preferable to accept that electron and the relativity theory, and regard it as a confirmation of the present theory of fine structure.

209. Experimental difficulties. There are considerable difficulties attaching to the size of the emitting atoms when the quantum numbers become large. If, for example, we confine ourselves to Bohr's original circular orbits, in the case of a Hydrogen atom, the radius in any stationary state is proportional to τ^2, where the number of quanta of angular momentum is τ—or the angular momentum of the electron is $\tau h/2\pi$. The "normal" radius of the Hydrogen atom ($\tau = 1$) on this theory is, for a circular orbit, of order 3.10^{-9}. In the solar chromosphere, 33 lines of the Balmer series are actually capable of measurement (Dyson), and the existence of the last one requires a sufficient proportion of the atoms of Hydrogen to be 33^2 or 1100 times the normal radius.

In vacuum tubes, the usual number of lines observed is 6 or 7, requiring a much smaller maximum atomic radius, and in fact, in the usual vacuum tube with a pressure of the order of 1 to 3 millimetres of mercury, such atoms have a free path, as calculation readily shows, quite consistent with the power of radiating as individual and undisturbed systems, without appreciable influence from neighbouring atoms. Larger atoms of the size required above are presumed capable of existence in the chromosphere on account of its low density, though the relation of this necessary low density to that indicated by other phenomena has not been discussed.

Recently, the problem has taken on a new aspect, for Merton* obtained 14 members of the Balmer series in a Helium tube at the great pressure of 42 millimetres of mercury. The Hydrogen in the tube was only the small quantity emitted from the electrodes when the discharge was passed, and all the lines were very sharp—a fact in itself probably indicating absence of disturbance in the electronic orbits on account of the proximity of neighbouring atoms.

Calculation shows that at this pressure, the necessary radius of the Hydrogen atom with $\tau = 14$ is actually greater than the mean distance apart of the Helium atoms. It would seem that this result contains valuable implications with regard to the "bound electron." For bound electrons in different atoms do not seem to influence each other, even

* *Proc. Roy. Soc.* (A), 1920.

at distances apart comparable with the radius, except perhaps catastrophically with interchanges of energy only in quanta, or in amounts which can permit the atoms to jump to new stationary states. But we shall not further discuss this aspect of the theory.

210. The general spectrum formula. We now proceed to indicate the manner in which the quantum theory of spectra must lead, in cases more complex than Hydrogen and charged Helium atoms, to the Ritz formula for spectral series, as a third approximation—though we sketch rather than give a full account of the investigation.

If a neutral atom loses one electron, this electron, on its return, is under the influence of a complex system of charge $+e$, and when it is sufficiently far away, the effect on the electron is that of a Hydrogen nucleus. Divergences from this effect occur because the field of potential in which the electron moves is not proportional to r^{-1} as in the previous theory. Actually, the field is not stationary, on account of the orbital motions in the rest of the atom—or the quasi-nucleus—but these motions are so rapid that a mean value may be used for the field.

If the atom has lost ν electrons, and one is returning, similar considerations are valid, and the field of potential is, in its first approximation,

$$V = +\frac{\nu e}{r}.$$

We are thus led to the consideration of the motion of an electron in a convergent mean field of potential of the type

$$V = \frac{\alpha_1}{r} + \frac{\alpha_2}{r^2} + \frac{\alpha_3}{r^3} + \dots,$$

where the α's rapidly decrease. Ordinarily the α's will be of orders proportional to the corresponding power of some length a determined by the nucleus and inner electrons, and which is very small compared with r, the distance of the electron from the origin.

In spherical polar coordinates, the equation of energy of the electron is

$$\tfrac{1}{2}m\{\dot{r}^2 + r^2\dot{\theta}^2 + r^2\sin^2\theta\dot{\phi}^2\} - e\left\{\frac{\alpha_1}{r} + \frac{\alpha_2}{r^2} + \dots\right\} = -W.$$

More strictly, the energy of the whole atom should be involved. We are neglecting the action of the returning electron on the rest of the atom, which is small, and even smaller when a mean value is taken for the atomic motions.

The momenta are

$$p_1 = m\dot{r}, \quad p_2 = mr^2\dot{\theta}, \quad p_3 = mr^2 \sin^2\theta\dot{\phi},$$

and as usual

$$p_3 = \text{constant} = n_3 h/2\pi,$$

$$p_2^2 + \frac{p_3^2}{\sin^2\theta} = \beta^2,$$

and therefore

$$\tfrac{1}{2}m\left(p_1^2 + \frac{\beta^2}{r^2}\right) - e\left(\frac{\alpha_1}{r} + \frac{\alpha_2}{r^2} + \dots\right) = -W,$$

while from the phase integral for p_2, we have as before

$$\beta = (n_2 + n_3)\, h/2\pi,$$

according to the additive property already described. The equation for p_1 is

$$p_1 = \sqrt{\left\{-2mW + \frac{2me\alpha_1}{r} - \frac{\beta^2 - 2me\alpha_2}{r^2} + 2me\sum_3^\infty \frac{\alpha_n}{r^n}\right\}}.$$

We must further suppose that $\beta^2 > 2me\alpha_2$, which, as can be seen readily from a consideration of orders of magnitude, will occur under our suppositions above. In actual fact, α_2 will ordinarily be zero, when mean values of the potential of such a system are taken, and only odd powers of r^{-1} will occur. We may suppose this to be the case without real loss of generality. Thus

$$n_1 h = \int dr \sqrt{\left\{-2mW + 2m\nu\frac{e^2}{r} - \frac{\beta^2}{r^2} + 2me\left(\frac{\alpha_3}{r^3} + \frac{\alpha_5}{r^5} + \dots\right)\right\}}.$$

If we neglect α_3 and later coefficients, this leads, of course, to previous analysis, with the result

$$W = \frac{2\pi^2 me^4\nu^2}{h^2} \cdot \frac{1}{(n_1 + n_2 + n_3)^2},$$

since

$$\beta = (n_2 + n_3)\, h/2\pi.$$

Taking account now of α_3, if

$$N = \sqrt{\left\{-2mW + 2m\nu\frac{e^2}{r} - \frac{\beta^2}{r^2}\right\}},$$

we may write, as the next approximation,

$$n_1 h = \int dr\left\{N + \frac{me\alpha_3}{r^3 N}\right\},$$

which can be taken twice between the critical values of r, making

$$N = 0.$$

We do not give the analysis, which presents no difficulty, and quote only the value of the integral, which becomes

$$n_1 h = 2\pi\left\{\frac{m\nu e^2}{\sqrt{2mW}} - \beta + \frac{m^2\nu e^3}{2\beta^3}\alpha_3\right\},$$

when α_3^2 is neglected. The value of W is

$$W = \frac{2\pi^2 me^4 \nu^2}{h^2} \left\{ n_1 + n_2 + n_3 - \frac{8\pi^4 m^2 \nu e^3 \alpha_3}{h^4 (n_2 + n_3)^3} \right\}^{-2}$$

For a system of given sum of angular quantum numbers, this is of the form

$$W = \frac{2\pi^2 me^4 \nu^2}{h^2} \cdot \frac{1}{(n_1 + \mu)^2},$$

where μ is constant. This leads to spectra of the Rydberg type, which naturally appear as second approximations in sequence to the Balmer type. They will obviously possess the more important practical limitations attached to the Rydberg formula, which is known to fail more and more as we approach the earlier members of the series. On the present view, this is a consequence of the smallness of n_1, the radial quantum number, for the earlier lines. The distance r of the emitting electron is not large in comparison with the atomic radius, and our neglect of α_5, α_3^2, ... is no longer justified.

Proceeding to the next order of approximation, we note that α_3 is of order a^2 in the length a, and α_5 of order a^4, so that α_5 and α_3^2 are of the same order and must be retained together.

The necessary integral is given, to the requisite order—using the binomial expansion—by

$$n_1 h = \int dr\, N \left\{ 1 + \frac{me\alpha_3}{N^2 r^3} + \frac{me\alpha_5}{N^2 r^5} - \frac{m^2 e^2 \alpha_3^2}{r^6 N^4} \right\}.$$

The evaluation of this integral—again between the roots of $N = 0$ twice—is somewhat tedious, and the analysis must be left to the reader.

The result in full is

$$\begin{aligned} n_1 h = 2\pi & \left\{ \frac{m\nu e^2}{\sqrt{2mW}} - \beta + \frac{m^2 e^3 \nu \alpha_3}{\beta^3} \right\} \\ & + \frac{\pi m \nu e^2}{\beta^3} \left\{ \frac{5m^3 e^5 \nu^2 \alpha_5}{2\beta^4} + \frac{15 m^3 \nu e^4 \alpha_3^2}{8\beta^4} \right\} \\ & - \frac{3\pi m W}{\beta^5} \left\{ \nu m^2 e^2 \alpha_5 + \tfrac{1}{4} m^2 e^2 \alpha_3^2 \right\}, \end{aligned}$$

to the order in question, namely a^4.

This leads to the form

$$\begin{aligned} W = \frac{2\pi^2 me^4 \nu^2}{h^2} \Bigg/ & \left\{ n_1 + n_2 + n_3 - \frac{8\pi^4 m^2 e^3 \nu \alpha_3}{h^4 (n_2 + n_3)^3} \right. \\ & \left. - \frac{80\pi^8 m^4 e^6 \nu^2}{h^8 (n_2 + n_3)^7} (3\alpha_3^2 + 4\nu e \alpha_5) + \frac{96 m^2 e^2 \pi^6}{h^6 (n_2 + n_3)^5} mW \left(\tfrac{1}{4}\alpha_3^2 + \nu e \alpha_5 \right) \right\}^2. \end{aligned}$$

For a given sum of angular quantum numbers, this depends on n_1 according to the law,—N being the Rydberg constant,—

$$\frac{W}{h} = N\nu^2 \bigg/ \left\{ n_1 + \mu_1 + \mu_2 \frac{W}{h} \right\}^2$$

where μ_1 and μ_2 are constants. This leads precisely to the Ritz law of spectra, according to which the "variable part" of a series of wave numbers n_1 is

$$\frac{N}{(n + \mu_1 + \mu_2 n_1)^2}$$

or, with similar approximation,

$$\frac{N}{\left[n + \mu_1 + \frac{\mu_2}{(n + \mu_1)^2} \right]^2}.$$

It is very remarkable that successive approximations to the quantum theory should give in turn (1) the Balmer form, (2) the Rydberg form and (3) the Ritz form of series, just in the historical order in which they were successively developed as measurements of spectra became more exact.

We do not continue the approximation further. The factor ν^2 is in accord with "Enhanced" or Spark spectra, and the theory of their origin from atoms which have lost more than one electron has been fortified very much recently by the experimental work of Fowler, notably in relation to the spectrum of magnesium.

The general analysis of phase-integrals, in fact, although we cannot give a full account of it in this treatise, makes it quite clear that in all arc spectra in the visual ranges for which the Rydberg constant is N, only one electron is primarily concerned in the emission, and that its orbits are described under an influence which, at least as a first approximation, is not very different from that of a Hydrogen nucleus. In other words, a single electron is proceeding, by jumps from one stationary state to another, towards a singly charged atom. In all spark spectra of constant "$4N$," the atom is doubly charged, and only one electron is returning. There is little doubt that the "super-spark" spectra, known in such elements as Carbon and Silicon, which are not yet investigated sufficiently fully to be arranged in series, will require constants $3^2 N$, $4^2 N$ for the constants of their series. Some of these super-spark spectra can be predicted at the present stage of the theory.

For example, we can determine the wave-lengths of the only possible spark spectrum of Lithium, on the supposition that its atomic number is 3. If it loses two electrons, thus retaining only one, its spectrum should be, in wave numbers,

$$n = 9N \left\{ \frac{1}{\tau_1^2} - \frac{1}{\tau_2^2} \right\}$$

with $N = 109730$ for Lithium. The only values in the range $\lambda\, 3000$ to $\lambda\, 7000$ are $n = 22220{\cdot}3,\ 19348{\cdot}3,\ 24072{\cdot}0$

with wave-lengths $\lambda = 4500{\cdot}4,\ 5168{\cdot}4,\ 4154{\cdot}2$

where only the first should be strong. In the same way, a super spark spectrum of Beryllium, with constant $16N$, can be calculated, but the spark spectra of these elements have not been investigated in the laboratory.

The remarks just made regarding the origin of N, $4N$, ... series must be applicable to Helium. These considerations amount to a demonstration, on the quantum basis, that the Helium atom, at least in the intervals of its radiation, exists in stationary states in which one electron is very much "internal" to the other, and the search for these states is the only path to a precise theoretical deduction of the ordinary Helium spectrum. Many model atoms for Helium have been proposed, but all fail to produce the known spectrum, though they are often capable of quantitative agreement with other phenomena. Perhaps one of the most interesting is that of Langmuir. In this model the electrons oscillate in step from P to Q, and R to S, and back again, the line joining them being always perpendicular to PQ and RS. The whole atom may also be rotating round its axis. This atom may be made to give the experimental value of the ionizing potential almost exactly, but it cannot give a spectrum depending on the simple Rydberg constant in the ordinary way.

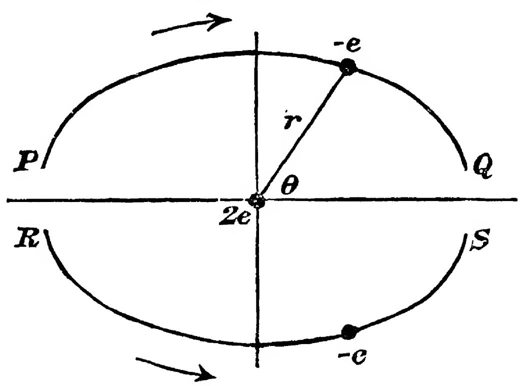


Fig. 188.

The current view is that the two electrons describe orbits of very different dimensions,—that motions are possible with the orbits nearly coplanar and also nearly perpendicular,—thus accounting for the two different types of series in the spectrum, consisting of sets of doublets and of single lines. The details of these arrangements have not been elucidated analytically, though much evidence has been adduced in favour of them by various writers.

211. Ionizing potentials. An important mode of investigation of the atom models of the quantum theory, while not strictly relevant to our point of view in this treatise, must be mentioned. A question which has been much disputed centres round the ionization potential of a Hydrogen atom, whose normal state is defined by

$$W = \frac{2\pi^2 m e^4}{h^2 \tau^2}$$

with $\tau = 1$.

If V is the potential of the surrounding field, measured in volts, the energy given to an electron is $\frac{1}{2}mv^2$, or

$$\frac{eV}{c}\,.\,10^8$$

so that if
$$\frac{eV}{c}\,.\,10^8 = W = \frac{2\pi^2 me^4}{h^2} = heN$$

where N is Rydberg's constant, the electron of the atom should be driven to infinity, and the atom ionized. Various experimental values for V have been found by different investigators. The theoretical value is

$$V = \frac{hc^2}{e}\,.\,N\,.\,10^{-8},$$

which we may evaluate as accurately as the data allow. The calculations of Sommerfeld and Bohr are well known. With the best measurements,

$$h = 6{\cdot}55\,.\,10^{-27},\ \ N = 1{\cdot}097\,.\,10^5,\ \ e = 4{\cdot}77\,.\,10^{-10},$$

and we find $V = 13{\cdot}56$ volts.

A quarter of this amount would ionize the atom in its second state $\tau = 2$, but only a small number of atoms should be in this state. The most accurate measurement is apparently that of Horton and Miss Davies*, who find 14·4 volts with a possibility that the value is too high.

For other atoms, spectral evidence alone indicates that the ring-arrangements are not the forms which ordinarily appear as stationary states. The evidence given by the ionization potentials points in the same direction. In the case of Helium, a very curious situation has existed in regard to this potential. Rau's experimental investigation (Wurzburg, 1914) showed that Helium was certainly ionized at 80 volts, and was not ionized at 75 volts. Several later investigators have shown that an ionization occurs also at about 25 volts.

A very recent paper by Compton and Lilly† appears to settle the question in favour of the quantum theory, although the authors only register details of results.

Let us consider the potentials to be expected in various circumstances. The energy of the ring arrangement in its normal state is

$$W = \frac{2\pi^2 me^4}{h^2}\left(2 - \frac{1}{4}\right)^2.\,2,$$

taking account of both electrons. This becomes

$$W = \frac{49}{8}\,heN,$$

* *Roy. Soc. Proc.* 1920.

† *Astrophys. Journal*, July, 1920.

and the ensuing potential is given by

$$\frac{eV}{c} \cdot 10^{-8} = \frac{49}{8} heN,$$

in order to drive both electrons away simultaneously. The value is

$$V = 83{\cdot}4 \text{ volts.}$$

If the second electron remains in the atom, the energy left in the normal state is

$$W = \frac{8\pi^2 me^4}{h^2},$$

and the potential required for the difference is 28·9 volts. The other 54·5 volts is required to ionize the charged atom. The ionizing potential of the ring atom is dependent on the orbit to which the remaining electron belongs. If it went to its fourth orbit, the potential would be almost precisely 80 volts, but there is no reason why this particular orbit should be chosen.

The experiments of Compton and Lilly are very exhaustive, for, unlike other investigators, they have found the potential for the various individual *types* of Helium spectrum, *i.e.*, Helium, Parhelium, the band spectrum, and the line λ 4686 due to the charged atom. They caused the emission by bombarding very pure Helium with the electrons from a hot filament cathode at various pressures, the limit being 24 mm. of Mercury. All the spectra appeared at 25·5 volts with low pressures and comparatively low current densities,—which are the conditions necessary for the present test, in which multiple impacts must be avoided. This must be regarded as the true ionizing potential of the Helium atom, but we can draw the further conclusion that the Helium and Parhelium spectra come from different types of states of the same atom. Moreover, the theoretical result is too high, and this experimental result, though not decisive, is adverse to the ring system.

The Pickering series has two ionizing potentials, 80 volts and 55 volts. It is clear that these represent the potentials necessary (i) to drive off both electrons at once, and (ii) to drive off a single electron from the charged atom. The second is precisely the theoretical value, and if the nature of the charged atom were not already sufficiently clear from the fine structure of its lines and other phenomena, this would be a strong confirmation, indicating also that these potentials can be measured to within about 2 %. But the same discrepancy occurs at 80 volts, the theoretical value for a ring being 83·4, as occurred at the 25·5 volts against 28·9. It seems to indicate that the true normal energy of a Helium atom is smaller than that of the ring system in the ratio

$$\frac{80}{83{\cdot}4} = 0{\cdot}959.$$

This fact would in itself, according to the principles of stability valid in the quantum theory, make the arrangement more probable, as a normal one, than the ring-system.

The fact that the band spectrum is produced at 25·5 volts is especially interesting. For it indicates that this spectrum, which cannot be attributed to a charged atom, is due to something which comes into existence at once in the presence of charged atoms. The only possibility seems to be a molecule, and it seems necessary to suppose that a Helium molecule He_2 with a positive charge can have a purely temporary existence under the conditions of discharge.

It should be noticed that the values of W which are possible in the various stationary states are known from the partial quotients in the spectra, even though the actual orbits in the atom may be beyond our knowledge. From these quotients, moreover, we can calculate the ionization potentials. For example, Hicks' formula for the wave numbers of the Helium Principal Series is

$$n = 38453{\cdot}35 - 109666{\cdot}2 \Big/ \left\{m + 0{\cdot}9294 + \frac{0{\cdot}0078}{m}\right\}^2$$

and if this is of the form predicted by quantum theory, putting $m = 1$, a possible ionizing potential is

$$V = 10^8 c^2 \frac{h}{e} \cdot \frac{109666{\cdot}2}{(1{\cdot}9372)^2},$$

which is fairly close to 80 volts.

212. Ehrenfest's principle. In proceeding to some of the more general considerations which belong to the optical side of the quantum theory, we shall begin with Ehrenfest's principle, which occupies, in some sense, a similar position to that once held by the "quasi-stationary" principle in the mechanics of the electron. It has been called the principle of "mechanical transformability" of stationary states. We do not propose to give any detailed analysis, for which reference must be made to original memoirs.

We suppose a system, with stationary states, in whose neighbourhood a field, say of electric or magnetic force, is slowly created and increased, and proceed to calculate the ensuing variation of the stationary states. Clearly it is not in fact to be calculated by ordinary dynamics, but nevertheless, if the field varies slowly enough, the states at any instant cannot differ widely from those corresponding to the instantaneous conditions of the field regarded as having been in operation permanently. It should thus be possible by ordinary dynamics to calculate the variations of states at least with good approximations when the field varies slowly. This is Ehrenfest's principle, called by him the "Adiabatic Hypothesis." It shows the possibility of a continuous mechanical connexion between two states, without which the equation

$$h\nu = W_1 - W_2,$$

has no meaning, for we cannot define the energies of two states unless there is some continuous connexion from one to the other by mechanical laws.

Considerable applications of this and allied conceptions have been made by Bohr recently. They have not led to new results in relation to spectra, but they have supplied comparatively simple proofs of the first approximations to such phenomena as the fine structure of Hydrogen lines, the Stark effect, and the Zeeman effect,—the two latter when the exciting magnetic fields are very small. The main function of the principle is in fact the simplification of the analysis in the deduction of first approximations.

With our formulation of the quantum theory, it is possible to give a simple demonstration of this principle at least when we deal only with a strictly periodic system with one degree of freedom. The analysis may be compared with Ehrenfest's which is of course similar, and Bohr has recently given a proof which is essentially the same, though looked at from what is really the reverse point of view, Bohr's object being to *demonstrate* the formula

$$\int p\,dq = nh,$$

as an invariant relation during the slow establishment of the field.

If a system has coordinates $q_1 \dots q_n$ and momenta $h_1 \dots h_n$, and an energy function E, kinetic energy T, and is in periodic motion with period τ, the integral

$$J = \sum_1^n \int_0^\tau p_r dq_r$$

$$= \int_0^\tau \sum_1^n p_r \dot{q}_r dt = 2\int_0^\tau T dt.$$

If there is another periodic motion, slightly different, under the influence of certain external forces, and if the variation from one motion to the other is denoted by δ,

$$\delta J = \int_0^\tau \sum_1^n (\dot{q}_r \delta p_r + \dot{p}_r \delta q_r)\, dt + \Big[\sum p_r \dot{q}_r \delta t\Big]_0^\tau$$

the square bracket being due to the change of period in the new motion. This may be written in the form

$$\delta J = \int_0^\tau \sum (\dot{q}_r \delta p_r - \dot{p}_r \delta q_r)\, dt + \Big[\sum p_r (\dot{q}_r \delta t + \delta q_r)\Big]_0^\tau$$

and the square bracket is zero since the new motion is periodic, the quantity in the bracket repeating itself. Hamilton's equations give

$$\dot{q}_r = \frac{\partial E}{\partial p_r}, \quad \dot{p}_r = -\frac{\partial E}{\partial q_r},$$

$$\Sigma\, \dot{q}_r \delta p_r - \dot{p}_r \delta q_r = \delta E,$$

and thus

$$\delta J = \int_0^T \delta E dt.$$

If the field has a potential V at any instant, and has required a long interval t_1 for its establishment to the value in the second motion, it began at $t = -t_1$, and has increased uniformly to the present instant. Then δE is the work it has done on all the charges of the system, so that

$$\delta E = -\int_{t=-t_1}^{t=0} \Sigma \frac{\partial V}{\partial q_r} \dot{q}_r dt \,.\, \frac{t_1 + t}{t_1}$$

(at $t = 0$), the factor $(t_1 + t)/t_1$ expressing the uniform increase. At a time t greater than zero, we add

$$-\sum_1^n \int_0^t \frac{\partial V}{\partial q_r} \dot{q}_r dt.$$

If we neglect the square of the field, values of the q's can be used corresponding to the first motion. We have also

$$\delta E = \frac{1}{t_1} \int_{-t_1}^0 V dt - (V),$$

where (V) is the value at a time t subsequently.

The first term is the mean of the second through a period, and so the integral of δE over a period is zero.

Thus $\delta I = 0$, or $\Sigma \int pdq$ is invariable for a system periodic at any moment during the establishment of the field. For one degree of freedom

$$\int pdq = nh$$

in the presence of the field, whose intensity must remain small.

The above contains Bohr's view of the necessary conditions involved in the presence of an external field, and he has not generalized it to larger fields. According to our view, this generalization is always possible, for any field could be replaced by a suitable number of charges in suitable positions, which can be regarded as an integral part of the system. The formulation

$$\int p_r dq_r = nh$$

should continue to be valid, and, for reasons outlined in an earlier section, it is preferable to lay down this relation as a unique and sufficient foundation for the whole theory.

213. Conditionally periodic systems, and separation of variables. The successful applications of the quantum theory have hitherto been entirely restricted to cases in which Jacobi's theorem allows us to separate the variables in the energy equation. We must at this point consider the underlying principles involved in the procedure, and we may begin with a statement of Jacobi's fundamental solution.

Let a system of m degrees of freedom be defined by the usual coordinates $q_1 \dots q_m$, and momenta $p_1 \dots p_m$, with an energy function E.

Then we have the usual set of equations

$$\dot{p}_r = -\partial E/\partial q_r, \quad \dot{q}_r = \partial E/\partial p_r,$$

when E is determinate in the p's and q's.

If there exists a function S, determined by the q's, so that

$$p_r = \partial S/\partial q_r,$$

the energy equation is of the form

$$E\left(q_1 \dots q_m \frac{\partial S}{\partial q_1} \dots \frac{\partial S}{\partial q_m}\right) = -W$$

where W is constant. There are also $m-1$ constants of integration in its complete solution.

It may happen that orthogonal coordinates are available, say the q's above, which allow a solution of the form

$$S = \Sigma S_r + C$$

where S_r is a function only of q_r, containing of course any number of constants. Then the system is said to allow "separation of variables." The momentum p_r then depends only on q_r, thus

$$p_r = \frac{\partial S}{\partial q_r} = \frac{\partial S_r}{\partial q_r}.$$

Moreover, p_r is the *square root* of a function of q_r, from the fact that mechanics gives E as containing the momenta as a sum of squares,—this is also true in the case of the mechanics of relativity, which is the reason for success of the theory in giving an account of the fine structure of Hydrogen and charged Helium lines. If $S_r = \sqrt{f(q_r)}$ we can write, accordingly,

$$S = \sum_1^m \int \sqrt{f_r(q_r)}\, dq_r,$$

each function f_r containing some or all of the integration constants. This is the sum, in essential, of a set of phase integrals. For if α_r, β_r

are the roots of $f_r(q_r) = 0$, the function S may be said to have a modulus of periodicity for every coordinate, the typical one being

$$J_r = \int \sqrt{f_r(q_r)}\, dq_r$$

taken from α_r to β_r and back again.

Theoretically, equations of this type can be reversed, and the various constants of integration, say $(\gamma_1, \gamma_2, \ldots)$ expressed as functions of the J's, say

$$\gamma_r = F_r(J_1 \ldots J_m)$$

and substituting for them, we have

$$\begin{aligned} S &= \Sigma S_r(q_r, J_1 \ldots J_m) \\ &= \Sigma \int \sqrt{f_r(q_r, J_1 \ldots J_m)}\, dq_r. \end{aligned}$$

According to our formulation, all the J's are of the form nh, where n is an integer. They are all determinate, with separation of variables, each p_r being expressed in terms of the corresponding q_r only.

214. Trigonometric series. Regarding the q's and p's as a set of variables at any moment, we can write

$$p_r = \frac{\partial S}{\partial q_r}, \quad u_r = \frac{\partial S}{\partial J_r}$$

where S is expressed entirely in q's and J's as in the formula last written. Then the p's and q's are transformed, and we have the system expressed in terms of the J's and u's.

Evidently if q_r goes through a complete cycle (from α_r to β_r and back again), u_r, regarded as dependent on the q's and J's, will increase by unity. But if one of the other q's goes through a cycle, u_r also goes through a cycle, and its value is not altered. Thus the q's and p's are periodic in the u's, with unit period, and they must admit expansions of the type

$$q_r = \Sigma (A_r)_{n_1 \ldots n_m} e^{2\pi i (n_1 u_1 + \ldots + n_m u_m)}$$

and the coefficients of this expansion are functions only of the J's,— which in our specification, means that they depend only on the quantum numbers determining the J's.

We may now introduce a theorem of Jacobi, to the effect that a transformation of the type we have made does not alter the canonical form of the equations of motion.

Thus they become, in the new variables

$$\dot{J}_r = -\frac{\partial E}{\partial u_r}, \quad \dot{u}_r = \frac{\partial E}{\partial J_r}.$$

But as E is the constant $-W$, $\partial E/\partial u_r = 0$, and thus

$$J_r = \text{constant}$$

this being the constant we take as of type nh.

Moreover $$u_r = \left(\frac{\partial E}{\partial J_r}\right) . t + \text{constant}$$
where $\frac{\partial E}{\partial J_r}$ is of the nature of generalized angular velocity.

The u's are accordingly called "angular coordinates." In fact, $\frac{\partial E}{\partial u_r}$ is the mean number of oscillations of q_r in unit time. Each coordinate q_r oscillates between limits dependent on the J's, and systems of this type are called "conditionally periodic." The displacement at any time, of a particle of the mechanical system, must be expressible uniquely in terms of the q's ; and must also therefore admit a trigonometric series of the same type as the series mentioned for the q's, the coefficients again depending only on the J's,—*i.e.* on quantum integers. Since
$$u_r = \frac{\partial E}{\partial J_r} . t + \text{const.} = \omega_r t + \text{const. (say)}$$
the frequencies of the harmonic analysis of any displacement are of type
$$n_1\omega_1 + n_2\omega_2 + \ldots + n_m\omega_m$$
where the n's are integers.

The paths of all particles are determinate in general—this point has been a source of considerable difficulty in the quantum theory at many points, for it has often been possible to choose the suitable orthogonal coordinates in more than one way—with, as a result, the same value of W in all cases, but totally different types of paths for the particles. The source of trouble lies in "degenerate" cases, in which motion does not occur in all the degrees of freedom.

The result above shows that the paths are strictly determinate if the number of these quantities ω is equal to the whole number of degrees of freedom, and if they are all mutually independent, with no identical relation between them.

We have not included, in our account, an illustrative case of such a degenerate system as is here mentioned, involving a comparative treatment by two different types of coordinates, as it did not appear to be necessary, and a statement of the existence of entirely distinct paths, but the same final W in the two cases, appears to be sufficient, and of course, it is the unique value of W which is important for spectral series, though the nature of the path-difficulty required some elucidation. A typical instance of a degenerate case is of course the Hydrogen atom undisturbed by any influence external to it. An alternative treatment for it, giving very different paths, is provided by the use of elliptic coordinates,—which readily give the solution, in place of spherical polars.

That there is, however, a unique solution of paths for these degenerate cases is now clear, if there is a unique solution for a conditionally periodic system of which they are special cases. The difficulty lies solely in its direct determination for the special case, and not as a particular type of a solution already obtained in general. Provided at least that they are degenerate cases of a conditionally periodic system,—in our view, of a system for which unique and finite phase integrals exist for each coordinate involved,—they must, as we have proved implicitly here, admit a definite solution.

It is of interest to show how precise Fourier series may be found for the representation of the motion. The method is essentially the usual astronomical method, though generalized in certain respects which it is not important to particularize.

Suppose that we wish to expand any function f of the q's of the system, the form being

$$f = \Sigma\Sigma \ldots A_{n_1, \ldots n_m} e^{2\pi\iota(n_1 v_1 + \ldots + n_m v_m)}.$$

Fourier's theorem gives the coefficients at once as

$$A_{n_1, \ldots n_m} = \int_0^1 \int_0^1 f \,.\, e^{-2\pi\iota(n_1 v_1 \ldots + n_m v_m)} \; dv_1 \, dv_2 \ldots dv_m$$

where f, under the integral sign, is regarded as being expressed in terms of the v's and J's, the coefficients A coming, as they should, to be functions only of the J's.

An integral with the q's as variables is obviously more convenient for analytical purposes, and this introduces the Jacobian D, or

$$D \equiv \frac{\partial(u)}{\partial(q)} \equiv \frac{\partial(u_1 \ldots u_m)}{\partial(q_1 \ldots q_m)}.$$

But, in terms of the q's, the u's are given by

$$u_r = \frac{\partial S}{\partial J_r} = \sum_{s=1}^{m} \frac{\partial S_s}{\partial J_r} = \sum_{s=1}^{m} \int \frac{\partial}{\partial J_r} \sqrt{f_s(q_s)}\, dq_s$$

and the determinant D is of a simple type, being the sum of products of functions, each containing only one q.

If f is of the same type as D then the final coefficient $A_{n_1 \ldots n_m}$ is of the type

$$\sum_r \phi_1^{(r)} \phi_2^{(r)} \ldots \phi_m^{(r)}$$

where the ϕ's are definite integrals of the form

$$\int \phi(q_r)\, e^{-2\pi\iota \sum_s n_s \frac{\partial S_r}{\partial J_s}} \, dq_r$$

ϕ being a function only of q_r, and the integral being between the limits of q_r and back again. They are in fact phase-integrals, and the problem is definitely solved analytically.

The trigonometric expansion is thus determinate, for the function f, under the hypothesis to which we have subjected f, namely that it is expressible as a set of terms each consisting of products of functions of single individual q's only. We see that this condition appears to be essential.

The function f ordinarily may be any Cartesian coordinate of any particle of the dynamical system, and in the case in which the q's belong to the wide class of elliptic or ellipsoidal coordinates (including parabolic, spherical, and so forth), it is known to be possible so to express the coordinates, as Jacobi showed*. No coordinates other than those included in this class have been effective hitherto in the quantum theory of spectra. In fact no others have been found in which the necessary separation of variables is possible, and it is believed but remains without proof, that no others exist. This belief has been expressed by H. A. Kramers, to whom reference should be made. The actual determination of any of these trigonometric series in a particular case is not included in this treatise. Many have been found by H. A. Kramers (Copenhagen).

They are used in obtaining prophecies of the intensity of component lines corresponding to passages between term-numbers. Other methods of determining them can be found, and their development is known in the case of a single Hydrogen atom, and, approximately, for the same atom disturbed by electric and magnetic fields. The results, regarding intensity of components, agree very well with observation, and seem to establish,—as worked out mainly by H. A. Kramers,—the foundation of this method of deducing intensities. But this discussion does not propose to enter into the question of the intensities of spectral lines in detail. Some great limitation of the scope of the subject was essential, and it was deemed desirable to concentrate attention upon the actual positions of lines in the spectrum, which must, on any view, constitute the final test of any theory purporting to give a genesis of spectra. For these reasons we shall only give a brief account of the Stark and Zeeman effect for hydrogen and charged Helium; at the same time a few remarks on what the authors consider to be the best point of view in the matter of polarization and intensity of lines seem to be called for. The problem is only considered in a very obscure and tentative manner by most writers who have dealt with it, and statements abound for which no justification can be seen readily. It seems possible, however, to put the essential part of the argument in a succinct form.

The first impetus was undoubtedly given by Einstein (*Phys. Zeit.* 18, 1917) when he developed his conception of the *à priori* probability of spontaneous transition between stationary states. This idea was

* *Vorles. uber Dyn.* p. 202.

actually introduced in connexion with temperature radiation. A system in a stationary state 1 must be supposed to possess an inherent probability of passing to any state 2 of smaller energy. Let this be called $A_{12}dt$, implying that the atom will perform the passage in time dt with this probability. This relates to an undisturbed atom. If it is surrounded by radiation, the presence of the radiation will introduce a new probability $B_{12}dt$, which must of necessity involve, as a factor, the density of the surrounding radiation. But in a vacuum tube in which the radiation is actually excited, it is clearly the probability of type A which is effective in determining the intensity of any spectral line. The atoms cannot be in temperature equilibrium with their surroundings, for the actual emission is due to bombardment by electrons which drive others out of the atom. Moreover, the density of the radiation cannot be large, a fact which in itself makes the effect of $B_{12}dt$ secondary. If n atoms are present in the tube originally, the energy contained in any radiation of frequency ν emitted in time dt must be nearly given by

$$n \,.\, h\nu \,.\, A_{12}dt,$$

if the atom emits one quantum at a time, ν being the frequency determined by passage between state 1 and state 2.

We have already mentioned Bohr's principle of analogy between the quantum process of radiation and the ordinary electrodynamics, which makes their results asymptotically equivalent in the region of long wave-lengths. This equivalence can be brought out very simply in the case of conditionally periodic systems—which for convenience, are supposed to be non-degenerate, so that the formulae of this section can be applied to them.

Let $n_1, n_2, \ldots, n_m$ be the integers applying to state 1 of such a system, and $n_1', n_n', \ldots, n_m'$ the corresponding integers for state 2. Any intermediate state can be defined by a J, such that

$$J_r = [n_1 + \lambda (n_1' - n_1)]\, h,$$

where λ can take any value from zero to unity, the states themselves appearing as the limits, and the intervening states all being "mechanically" possible. The frequency emitted during the transition between the two states is at once found to be the mean value, from $\lambda = 0$ to $\lambda = 1$, of the frequency

$$(n_1 - n_1')\,\omega_1 + (n_2 - n_2')\,\omega_2 + \ldots,$$

arising in the harmonic analysis of the electronic motion. Otherwise, we have, naturally

$$\delta E = \sum_1^m \omega_r \delta J_r,$$

between the two states, or its equivalent for frequency,

$$\nu = \int_{\lambda=0}^{1} \frac{\delta E}{h} = \int_{\lambda=0}^{1} \left(\frac{\omega_1}{h} \delta J_1 + \frac{\omega_2}{h} \delta J_2 + \dots \right) = \int_0^1 d\lambda \, \{(n_1 - n_1') \, \omega_1 + \dots\},$$

as stated.

The region of large wave-length signifies that for which the integers n are all large in the states concerned,—small wave numbers in the resulting lines of the spectrum,—and if $n_1 - n_1'$, $n_2 - n_2'$, ... are small, the ω's are practically unaltered from one state to the other, and we can replace the integral by

$$(n_1 - n_1') \, \omega_1 + \dots,$$

and emission follows the characteristic frequencies of systems themselves, as in electrodynamics.

From this correspondence in the large wave-length region, we can infer that other phenomena of lines will tend asymptotically in this region to the phenomena to be expected on ordinary dynamics. As an illustration, we already know that Planck's formula for radiation, which is, as shown by W. Wilson, a formal consequence of our supposition, tends to that of Rayleigh and Jeans in this region. We may suppose that such phenomena as polarization and intensity of spectral lines behave in a like manner.

Now for an electron performing the linear motion

$$x = A \cos 2\pi f t,$$

(frequency f), the radiation is proportional to $(\ddot{x})^2$ or to $A^2 f^4$. We deduce that the *à priori* probability of spontaneous transition from state 1 to state 2, when integers are large and

$$n_1' - n_1 = r, \quad n_2' - n_2 = r_2, \quad \dots,$$

is proportional to

$$A^2 f^4 / hf \quad \text{or} \quad A^2 f^3,$$

where

$$f = \Sigma \, \omega_r r_r,$$

is the frequency of the emitted radiation.

We may conclude that the radiation is polarized along one direction, also, and suppose a generalization by which, if an electron moves in ordinary dynamics in such a way as to produce other types of polarization, the same will apply asymptotically here, directions of the harmonic motions of the analysis replacing directions of motion of the analogous electron in ordinary dynamics. According to this basis, the coefficients in our harmonic analysis really determine, accurately for large wave-lengths, and at least roughly for small ones (or small integers) the intensities of emitted lines, by their squares. They determine at least partially the *à priori* probability of transition between two states,—we recall that the coefficients in question depend only on the J's, *i.e.* on the integers characterizing the states.

Such conclusions regarding polarization of components have been found effective in the discussion of Stark and Zeeman effects. The intensity problem is by no means in so satisfactory a condition. For we do not know the number of atoms involved, or how to make a suitable choice for the amplitude A, which is *some* mean value of amplitude over all the values possible in the intervening states characterized typically by the symbol λ which we introduced. The number of atoms involved varies in ways quite unknown with a large variety of types of experimental condition prevalent in vacuum tubes. Any intensity determination which can be predicted is thus very rough.

This criticism, or rather failure to solve the problem, does not, however, apply in all cases, and especially in the case of a degenerate system tested for Zeeman effect, Stark effect, or even fine-structure under the relativity conditions. For a system, initially degenerate, in the first place ceases to be so in the presence of a field. For example, a hydrogen atom whose motion is in one plane has its other angular coordinate brought into play when a field is applied. The field splits the lines of the atom into components, each with characteristic integers. The energies in state 1 are effectively equal, and we may suppose the number of atoms in any state to be proportional to the probability of that state. When this supposition is made, the results, for instance in the intensities of lines in the case of the Stark effect, appear to support it fairly well, so that an exception must be made, in this especial instance, to our conclusion that the general problem of intensity is one of extreme difficulty.

215. The Stark effect. In order to determine the nature of the stationary states of an atom under the influence of some external field, we must consider in what manner this field alters the decomposition of the motion into a set of harmonic oscillations. For the orbit is necessarily varying steadily in the presence of the field, and a set of harmonic components may not be capable of representing the motion at all. In other words, a certain definiteness in the stationary states will be lost, and the effect to be looked for is a broadening of the spectral lines of the undisturbed atom—though any atom, at the moment of its emission, is sending out a monochromatic radiation.

Yet cases will occur when the effect of the field is sufficiently simple to allow a resolution into harmonic components, for example, when the variations in the orbit are themselves periodic, and an orbit is repeated at regular intervals. Such intervals are periods of the orbital perturbations, and should, in harmonic analysis, give rise to periods which are multiples of them. If we apply, further, the principle of correspondence as used by Bohr, we see the possibility of a whole set of stationary states in the system with an external field, corresponding to each individual state in the undisturbed system. In a transition between two of these states, a radiation should be emitted whose relation to the frequency of variation of the orbit is the same as that which occurs in a simple periodic system in relation to its period.

Let us suppose, for instance, that a Hydrogen atom is situated in a uniform external electric field. The orbit of its electron varies continuously both as regards its eccentricity and its major axis,—we are regarding it, for simplicity, as moving in two dimensions in a slowly changing ellipse. But the centre of the orbit moves in a plane perpendicular to the field, and in a periodic manner. When it returns to any starting point, the eccentricity and other features of the orbit will recur, and the whole motion is then repeated. In this very simple case the period of repetition depends only on the original eccentricity (e) and major axis (a), and is given by

$$\sigma = \frac{3eF}{8\pi^2 ma\omega}$$

where F is the external force of the field, ω being the frequency of revolution in the ellipse.

The energy difference between two states corresponding to the same undisturbed state is a multiple of $h\sigma$ or $sh\sigma$, so that if E_U is the energy of the undisturbed state, that of any disturbed one associated with it is

$$E = E_U + sh\sigma.$$

Thus $$E = E_U + \frac{sh \, . \, 3eF}{8\pi^2 ma\omega}.$$

Now in the undisturbed state, if the total quantum number is n,

$$E_U = -W = -\frac{2\pi^2 me^4}{h^2 n^2}, \quad 2a = \frac{n^2 h^2}{2\pi^2 e^2 m}, \quad \omega = \frac{4\pi^2 me^4}{h^3 n^3},$$

and therefore $$E = -\frac{2\pi^2 me^4}{h^2 n^2} + ns \, . \, \frac{3h^2 F}{8\pi^2 me}.$$

The frequency of radiation emitted in passage between two disturbed states characterized by s_1, s_2, whose corresponding undisturbed states are characterized by (n_1, n_2) is therefore

$$\frac{2\pi^2 me^4}{h^3}\left\{\frac{1}{n_1^2} - \frac{1}{n_2^2}\right\} + \frac{3hF}{8\pi^2 me}(n's' - ns).$$

The undisturbed spectral line has, in the presence of the field, a variety of new components whose separations are determined by all the possible values of $n's' - ns$—where, as a necessary restriction, s is less than n, and s' less than n'. This formula is found experimentally to give a complete account of the Stark effect in Hydrogen.

The strict proof of the formula, given by Epstein and Schwarzschild, involves the use of parabolic coordinates when the field is present, and the evaluation of phase integrals by approximate methods. These coordinates allow the variables to be separated after the Jacobi manner in the equation of energy, but the investigation is very long and we do not reproduce it. In fact, it is the most general case of a conditionally periodic system which has been solved in this way.

The harmonic decomposition of the motion in the atom is of the form, where ξ is the displacement of the electron in any direction,

$$\xi = \Sigma\Sigma A_{r,s} \cos 2\pi \{(r\omega + s\sigma) t + a_{r,s}\}$$

where ω is the mean frequency in the orbit, and σ is the frequency of the perturbations. The summation is for integer values of r and s. When all the integers involved are large, the frequency of the line emitted in passage from (r, s) to (r', s') is approximately

$$(r - r')\,\omega + (s - s')\,\sigma.$$

The correspondence principle can be applied to give data regarding the polarization of the Stark components, but this aspect of the matter is hardly yet definite enough for detailed discussion. Reference may be made to a recent work by Bohr*, which contains a similar account of the Zeeman effect.

* *Theory of Spectra and Atomic Constitution*, Cambridge University Press, 1922.

INDEX

(The numbers indicate references to pages.)

CAMBRIDGE: PRINTED BY W. LEWIS AT THE UNIVERSITY PRESS

MAGNETISM AND ELECTRICITY. By J. PALEY YORKE, Head of the Physics and Electrical Engineering Department, L.C.C. School of Marine Engineering, Poplar. Crown 8vo, 5s.

AN INTRODUCTION TO PRACTICAL PHYSICS. By E. H. BARTON, D.Sc., F.R.S.E., Professor of Experimental Physics, University College, Nottingham; and the late T. P. BLACK, M.Sc., Ph.D. 55 Figures. SECOND EDITION. Crown 8vo, 5s.

ADVANCED EXAMPLES IN PHYSICS. By A. O. ALLEN, M.A., B.Sc., A.R.C.S., Lecturer in Optics at Leeds University. SECOND EDITION. Crown 8vo, 2s. 6d. net.

MATHEMATICS

FIVE-FIGURE TABLES OF MATHEMATICAL FUNCTIONS. Comprising Tables of Logarithms, Powers of Numbers, Trigonometric, Elliptic, and other Transcendental Functions. By J. B. DALE, M.A., Assistant Professor of Mathematics at King's College, London. Demy 8vo, 4s. 6d. net.

LOGARITHMIC AND TRIGONOMETRIC TABLES (To Five Places of Decimals). By J. B. DALE, M.A. Demy 8vo, 2s. 6d. net.

THE CALCULUS FOR ENGINEERS. By JOHN PERRY, M.E., D.Sc., F.R.S., late Emeritus Professor of Mechanics and Mathematics in the Royal College of Science, London. THIRTEENTH IMPRESSION. viii + 382 pages. Crown 8vo, 8s. 6d.

CALCULUS FOR SCHOOLS. By R. C. FAWDRY, M.A., Head of the Military and Engineering Side at Clifton College; and C. V. DURELL, Senior Mathematical Master at Winchester College. Crown 8vo. With Answers. In one volume, 6s. 6d. Part I, 3s. 6d.; Part II, 4s.

AN INTRODUCTION TO PROJECTIVE GEOMETRY. By L. N. G. FILON, M.A., D.Sc., F.R.S., Professor of Applied Mechanics, University College, University of London. THIRD EDITION. viii + 261 pages. Crown 8vo, 7s. 6d.

MATHEMATICAL DRAWING. Including the Graphic Solution of Equations. By the late G. M. MINCHIN, M.A., F.R.S.; and J. B. DALE, M.A. THIRD EDITION, SECOND IMPRESSION. vi + 145 pages. Demy 8vo, cloth, 8s. 6d. net.

HIGHER ALGEBRA. By W. P. MILNE, M.A., D.Sc., Professor of Mathematics in the University of Leeds. xii + 586 pages. Crown 8vo, 8s. 6d.

HOMOGENEOUS CO-ORDINATES. By W. P. MILNE, M.A., D.Sc. xii + 164 pages. Crown 8vo, 6s. net.

EXAMPLES IN THE MATHEMATICAL THEORY OF ELECTRICITY AND MAGNETISM. By J. G. LEATHEM, M.A., Fellow of St. John's College, Cambridge. Crown 8vo, cloth, 2s.

VECTORS AND ROTORS: with applications. By O. HENRICI, Ph.D., LL.D., F.R.S.; and G. C. TURNER, B.Sc., Head of the Mathematics and Physics Department, Goldsmiths' College. xvi + 204 pages. Crown 8vo, 5s.

GEOMETRICAL CONICS. By G. W. CAUNT, M.A., Lecturer in Mathematics, Armstrong College, Newcastle-on-Tyne; and C. M. JESSOP, M.A., Professor of Mathematics, Armstrong College, Newcastle-on-Tyne. Crown 8vo, 3s.

ENGINEERING

STRENGTH AND STRUCTURE OF STEEL AND OTHER METALS. By W. E. DALBY, F.R.S., M.A., B.Sc., M.Inst.C.E., M.I.M.E., University Professor of Engineering at the City and Guilds (Engineering) College. Very fully illustrated. 192 pages and 38 plates. 8vo, 18s. net.

STEAM POWER. By Professor W. E. DALBY, F.R.S., M.Inst.C.E., M.I.M.E. SECOND EDITION. xvi + 760 pages, with 250 diagrams. 8vo, 25s. net. The Total Energy and Temperature Diagram and the Total Energy and Entropy Diagram (Mollier Diagram) are obtainable separately, price 6d. net each (by post in a tube, 11d.).

VALVES AND VALVE GEAR MECHANISMS. By Professor W. E. DALBY, F.R.S. xviii + 366 pages, 202 illustrations. SECOND IMPRESSION. Royal 8vo, 24s. net.

THE BALANCING OF ENGINES. By Professor W. E. DALBY, F.R.S. THIRD EDITION. xii + 283 pages, 184 illustrations. Demy 8vo, 12s. 6d. net.

PROPERTIES OF STEAM AND THERMODYNAMIC THEORY OF TURBINES. By H. L. CALLENDAR, F.R.S., Professor of Physics in the Imperial College of Science and Technology. 544 pages, numerous diagrams. 8vo, 30s. net.

THE CALLENDAR STEAM TABLES. 3s. 6d. net.

ABRIDGED CALLENDAR STEAM TABLES. (Centigrade Units.) 8vo, 1s. net.

ABRIDGED CALLENDAR STEAM TABLES. (Fahrenheit Units.) 8vo, 1s. net. These abridged tables contain the necessary data in a convenient form for Engineers and Students engaged in steam work.

THE CALLENDAR STEAM DIAGRAM. (Centigrade Units.) This is supplied with "The Properties of Steam," and also with "The Callendar Steam Tables." It can be obtained separately, price 6d. net (by post in a tube, 11d.).

THE CALLENDAR STEAM DIAGRAM. (Fahrenheit Units.) This is *not* supplied with "The Properties of Steam" or with "The Callendar Steam Tables." 6d. net (by post in a tube, 11d.).

HEAT DROP TABLES: ABSOLUTE PRESSURES. Calculated by H. Moss, M.Sc., A.R.C.S., from the Formulæ and Steam Tables of Professor H. L. Callendar, F.R.S. 64 pages. Cloth, 5s. net.

HEAT DROP TABLES: H.P. GAUGE PRESSURES, L.P. ABSOLUTE PRESSURES. Calculated by H. Moss, Imperial College of Science, from Professor Callendar's Formulæ and Steam Tables. Cloth, 5s. net.

CORRECTION TABLES FOR THERMODYNAMIC EFFICIENCY. Calculated by C. H. NAYLOR, Assoc.M.Inst.C.E. Cloth, 5s. net.

STEAM TURBINE DESIGN. With Especial Reference to the Reaction Type, and including Chapters on Condensers and Propeller Design. By JOHN MORROW, M.Sc., D.Eng. (Armstrong College, Newcastle-on-Tyne).
[*New Edition in Preparation.*

RAILWAY ELECTRIC TRACTION. By F. W. CARTER, M.A., M.I.E.E., British Thomson-Houston Co., Rugby. viii + 412 pages, with 204 illustrations and 10 folding plates. Demy 8vo, 25s. net.

HYDRAULICS. For Engineers and Engineering Students. By F. C. LEA, D.Sc., M.Inst.C.E., Professor of Civil Engineering in the University of Birmingham. FOURTH EDITION. xii + 594 pages, 400 diagrams. Demy 8vo, 18s. net.

ELEMENTARY HYDRAULICS. For Technical Students. By F. C. LEA, D.Sc., M.Inst.C.E. viii + 224 pages, with 156 diagrams. Crown 8vo, 7s. 6d. net.

MODERN METHODS OF WATER PURIFICATION. By JOHN DON, F.I.C., A.M.I.Mech.E., and JOHN CHISHOLM, A.M.I.Mech.E. SECOND EDITION. xviii + 398 pages, 106 illustrations. Demy 8vo, 16s. net.

MODERN ROADS. By H. P. BOULNOIS, M.Inst.C.E., F.R.San.Inst., etc., Chairman of the Council of the Roads Improvement Association, etc. xii + 302 pages. Demy 8vo, 16s. net.

REINFORCED CONCRETE DESIGN. VOL. I.: THEORY. By OSCAR FABER, D.Sc., A.M.Inst.C.E., Chief Engineer to Messrs. Trollope and Colls, Ltd., and P. G. BOWIE, A.M.Inst.C.E. xx + 332 pages, 158 diagrams. SECOND EDITION. Demy 8vo, 14s. net. VOL. II.: PRACTICE. By OSCAR FABER, D.Sc., A.M.Inst.C.E. xii + 246 pages, 89 diagrams. Demy 8vo, 18s. net.

THE PRACTICE OF RAILWAY SURVEYING AND PERMANENT WAY WORK. By S. WRIGHT PERROTT, M.A.I., M.Inst.C.E., Professor of Civil Engineering in the University of Liverpool; and F. E. G. BADGER, A.M.Inst.C.E. viii + 304 pages, with 140 diagrams. Demy 8vo, 30s. net.

SURVEYING. By W. N. THOMAS, M.Sc. (Birmingham), B.Sc. Eng. (London), Assoc.M.Inst.C.E., A.M.I.Mech.E., A.M.Inst. M. and Cy.E., Demonstrator in Engineering at the University of Oxford. viii + 536 pages and 298 diagrams. 8vo, 25s. net.

THE FIELD ENGINEER'S HANDBOOK. By G. C. WELLS and A. S. CLAY, B.Sc. SECOND EDITION. viii + 227 pages, 71 illustrations. Small 8vo, 8s. 6d. net.

TRAVERSE TABLES. With an Introductory Chapter on Co-ordinate Surveying. By HENRY LOUIS, M.A., D.Sc., M.I.C.E., Professor of Mining; and G. W. CAUNT, M.A., Lecturer in Mathematics, Armstrong College, Newcastle-on-Tyne. SECOND EDITION. 8vo, 5s. 6d. net.

TACHEOMETER TABLES. By H. LOUIS, M.A., D.Sc., M.I.C.E.; and G. W. CAUNT, M.A. 8vo, 10s. 6d. net.

A TEXT-BOOK OF ELECTRICAL ENGINEERING. By Dr. A. THOMÄLEN. Translated by Professor G. W. O. HOWE, D.Sc. FOURTH EDITION. xii + 482 pages, 480 diagrams. Demy 8vo, 28s. net.

THE PRINCIPLES OF ELECTRICAL ENGINEERING AND THEIR APPLICATION. By Dr. G. KAPP, late Professor of Electrical Engineering in the University of Birmingham. VOLUME I.: PRINCIPLES. Fully illustrated. xii + 356 pages. Demy 8vo, 18s. net. VOLUME II.: APPLICATION. x + 388 pages. 18s. net.

THE FOUNDATIONS OF ALTERNATE CURRENT THEORY. By C. V. DRYSDALE, D.Sc. (London), M.I.E.E., F.R.S.E. xii + 300 pages, 190 illustrations. Demy 8vo, 10s. 6d. net.

THE THEORY OF MACHINES. By R. F. MCKAY, M.Sc., A.M.Inst. C.E., A.M.I.Mech.E. SECOND EDITION. viii + 440 pages, 407 diagrams. Demy 8vo, 20s. net.

GRINDING MACHINERY. By J. J. GUEST, M.A., M.I.Mech.E., Professor of Mechanical and Electrical Engineering, Artillery College, Woolwich. xii + 444 pages, with illustrations. Demy 8vo, 16s. net.

STEAM BOILERS AND BOILER ACCESSORIES. By the late W. INCHLEY, B.Sc. xi + 412 pages, 141 illustrations. Crown 8vo, 8s. 6d. net.

THE PRACTICAL DESIGN OF MOTOR-CARS. By J. GUNN. viii + 256 pages, 162 illustrations. Demy 8vo, 12s. 6d. net.

THE STRENGTH AND ELASTICITY OF STRUCTURAL MEMBERS. By R. J. WOODS, M.E., M.Inst.C.E. SECOND EDITION. xii + 310 pages, 292 illustrations. Demy 8vo, 14s. net.

THE THEORY OF STRUCTURES. By R. J. WOODS, M.E., M.Inst. C.E. xii + 276 pages, 157 illustrations. Demy 8vo, 12s. 6d. net.

THE ITALIAN ORDERS OF ARCHITECTURE. By CHARLES GOURLAY, B.Sc., A.R.I.B.A., Professor of Architecture and Building in the Royal Technical College, Glasgow. With 32 full-page plates. Large 4to (12" × 9½"), 6s. net.

MECHANICAL DRAWING. With Special Reference to the Needs of Mining Students. By JOSEPH HUSBAND, B.Eng., A.M.I.C.E., Professor of Civil Engineering at Sheffield University. With 40 plates. Quarto, 3s. 6d. net.

MACHINE SKETCHES AND DESIGNS. By Professor A. CRUICKSHANK, M.I.Mech.E., and R. F. MCKAY, M.Sc., A.M.Inst.C.E. viii + 40 pages of illustrations, with short Explanatory Notes. SECOND EDITION. Quarto, 2s. 6d. net.

PETROL ENGINE CONSTRUCTION AND DRAWING. By W. E. DOMMETT, Wh. Ex., A.M.I.A.E. 26 plates, with descriptive letterpress and questions for technical students. Demy 4to, 3s. 6d. net.

EXAMPLES IN ELECTRICAL ENGINEERING. By J. F. GILL, M.Sc., B.Eng., A.M.I.Mech.E.; and F. J. TEAGO, M.Sc., M.I.E.E., The University, Liverpool. Crown 8vo, 7s. 6d. net.

EXERCISES IN ELECTRICAL ENGINEERING. For the Use of "Second Year" Students in Universities and Technical Colleges. By T. MATHER, F.R.S., M.I.E.E., Professor of Electrical Engineering, City and Guilds (Engineering) College; and G. W. O. HOWE, D.Sc., M.I.E.E., Assistant Professor of Electrical Engineering, City and Guilds (Engineering) College. THIRD EDITION. Crown 8vo, 2s. net.

GEOLOGY AND MINING

THE GEOLOGY OF THE BRITISH EMPIRE. By F. R. C. REED, Sc.D., F.G.S., formerly Assistant to the Professor of Geology in the University of Cambridge. viii + 480 pages, with 25 maps and sections. Demy 8vo, 30s. net.

COAL IN GREAT BRITAIN. By WALCOT GIBSON, D.Sc., F.G.S., F.R.S.E., Head of the Geological Survey of Scotland. viii + 312 pages, with 50 diagrams and 8 plates. Demy 8vo, 21s. net.

OIL FINDING: An Introduction to the Geological Study of Petroleum. By E. H. CUNNINGHAM CRAIG, B.A., F.G.S. SECOND EDITION. SECOND IMPRESSION. xii + 324 pages, 13 plates and 20 illustrations. Demy 8vo, cloth, 16s. net.

THE DRESSING OF MINERALS. By H. LOUIS, D.Sc., M.I.M.E., M.I.C.E., Professor of Mining and Lecturer on Metallurgy at Armstrong College, Newcastle-on-Tyne. x + 544 pages, 416 illustrations. Super royal 8vo, cloth, 30s. net.

WINDING ENGINES AND WINDING APPLIANCES: Their Design and Economical Working. By G. MCCULLOCH, A.M.I.M.E., Inspector of Machinery in the Department of Mines, West Australia; and T. C. FUTERS, M.Inst.M.E. viii + 452 pages, 175 illustrations. Demy 8vo, 21s. net.

A TEXTBOOK OF GEOLOGY. By P. LAKE, M.A., F.G.S., Royal Geographical Society Reader in Regional and Physical Geography in the University of Cambridge; and R. H. RASTALL, Sc.D., F.G.S., Lecturer in Economic Geology in the University of Cambridge. xiv + 508 pages, fully illustrated. THIRD EDITION. Demy 8vo, 21s. net.

OUTLINES OF PALÆONTOLOGY. By H. H. SWINNERTON, D.Sc., F.G.S., Professor of Geology at University College, Nottingham. Fully Illustrated. Demy 8vo, cloth, 30s. net.

THE GEOLOGY OF ORE DEPOSITS. By H. H. THOMAS, M.A., B.Sc., and D. A. MACALISTER, Assoc. R.S.M., of the Geological Survey of Great Britain. xii + 416 pages and 65 illustrations. Crown 8vo, 8s. 6d. net.

THE GEOLOGY OF BUILDING STONES. By J. ALLEN HOWE, B.Sc., Curator of the Museum of Practical Geology. viii + 455 pages, fully illustrated. Crown 8vo, 8s. 6d. net.

THE GEOLOGY OF SOILS AND SUBSTRATA. By the late H. B. WOODWARD, F.R.S. xvi + 366 pages, with illustrations. Crown 8vo, 8s. 6d. net.

GEOLOGICAL AND TOPOGRAPHICAL MAPS: Their Interpretation and Use. By A. R. DWERRYHOUSE, D.Sc., F.G.S. SECOND EDITION. viii + 133 pages, with 90 illustrations. Demy 8vo, 6s. net.

AN INTRODUCTION TO MINE SURVEYING. By T. BRYSON, A.R.T.C., M.Inst.M.E., Lecturer in Mining at the Mining and Technical College, Wigan, and G. M. CHAMBERS, M.Inst.M.E., Certificated Colliery Manager and Surveyor of Mines. Fully illustrated. viii + 288 pages, 6s. net.

CHEMISTRY AND METALLURGY

METALS AND METALLIC COMPOUNDS. By U. R. EVANS, M.A., King's College, Cambridge. 4 vols., obtainable separately. Demy 8vo. Vol. I., 21s. net. Vol. II., 18s. net. Vol. III., 14s. net. Vol. IV., 18s. net.

ASSAYING IN THEORY AND PRACTICE. By E. A. WRAIGHT, A.R.S.M., M.I.M.M., late Senior Demonstrator in the Assaying Laboratory of the Royal School of Mines, London. xii + 324 pages, 70 illustrations. Demy 8vo, 10s. 6d. net.

SERVICE CHEMISTRY. By the late VIVIAN B. LEWES, F.I.C., F.C.S.; and J. S. S. BRAME, F.I.C., F.C.S., Professor of Chemistry, Royal Naval College, Greenwich. FIFTH EDITION. xvi + 576 pages. Illustrated. Demy 8vo, 21s.

FUEL. Solid, Liquid, and Gaseous. By J. S. S. BRAME, F.I.C. SECOND EDITION. xiv + 372 pages, with 73 illustrations and 55 tables. Demy 8vo, 15s. net.

PETROL AND PETROLEUM SPIRITS. A Description of their Sources, Preparation, Examination, and Uses. By W. E. GOODAY, A.R.S.M., D.I.C., A.M.Inst.P.T. With a Preface by Sir JOHN CADMAN, K.C.M.G. xii + 135 pages. Illustrated. Demy 8vo, 10s. 6d. net.

THE ABSORPTION OF NITROUS GASES. By H. W. WEBB, M.Sc., F.I.C., Head of Chemistry Department, Cardiff Technical College. Illustrated. Demy 8vo, 25s. net.

THE CHEMISTRY AND TESTING OF CEMENT. By C. H. DESCH, D.Sc., Ph.D., Professor of Metallurgy in the University of Sheffield. [*New Edition in Preparation.*

THE CHEMISTRY AND MANUFACTURE OF HYDROGEN. By P. LITHERLAND TEED, A.R.S.M., A.I.M.M., Major, R.A.F. Illustrated. Demy 8vo, cloth, 10s. 6d. net.

THE PRINCIPLES OF APPLIED ELECTRO-CHEMISTRY. By A. J. ALLMAND, D.Sc., Professor of Chemistry, King's College, London. xii + 547 pages, with 136 figures. Demy 8vo, 25s. net.

FORENSIC CHEMISTRY. By A. LUCAS, O.B.E., F.I.C., Director of the Government Analytical Laboratory and Assay Office, Cairo. viii + 268 pages. Demy 8vo, 15s. net.

ORGANIC CHEMISTRY FOR ADVANCED STUDENTS. By JULIUS B. COHEN, Ph.D., B.Sc., F.R.S., Professor of Organic Chemistry in the University of Leeds. FOURTH EDITION, in Three Parts. Each Part obtainable separately. Demy 8vo, cloth.

THE CHEMICAL SYNTHESIS OF VITAL PRODUCTS, AND THE INTER-RELATIONS BETWEEN ORGANIC COMPOUNDS. By the late Professor RAPHAEL MELDOLA, F.R.S., F.I.C., etc. xvi + 338 pages. Super royal 8vo, 21s. net.

BIO-CHEMISTRY. A Study of the Origin, Reactions, and Equilibria of Living Matter. By the late BENJAMIN MOORE, M.A., D.Sc., F.R.S., Whitley Professor of Bio-Chemistry in the University of Oxford. viii + 340 pages. Demy 8vo, 21s. net.

A LABORATORY HANDBOOK OF BIO-CHEMISTRY. By P. C. RAIMENT, B.A., M.R.C.S., and G. L. PESKETT, B.A., The University, Oxford. Crown 8vo, 5s. net.

CHEMICAL DISINFECTION AND STERILIZATION. By S. RIDEAL, D.Sc., F.I.C., and E. K. RIDEAL, M.A., D.Sc., F.I.C. viii + 313 pages. Demy 8vo, 21s. net.

LEAD POISONING AND LEAD ABSORPTION. By T. M. LEGGE, M.D., H.M. Medical Inspector of Factories; and K. W. GOADBY, M.R.C.S., D.P.H. 308 pages. Illustrated. Demy 8vo, 12s. 6d. net.

INDUSTRIAL POISONING. Translated from Dr. Rambousek's "Gewerbliche Vergiftungen" by Dr. T. M. LEGGE. Demy 8vo, 14s. net.

PHYSICAL CHEMISTRY: its Bearing on Biology and Medicine. By J. C. PHILIP, D.Sc., F.R.S., Professor of Physical Chemistry in the Imperial College of Science and Technology. SECOND EDITION. Crown 8vo, 8s. 6d. net.

ELEMENTARY PHYSICAL CHEMISTRY. By W. H. BARRETT, M.A., Harrow School. viii + 247 pages, with 61 diagrams. 6s.

THE CHEMISTRY OF COLLOIDS AND SOME TECHNICAL APPLICATIONS. By W. W. TAYLOR, M.A., D.Sc., Lecturer in Chemical Physiology in the University of Edinburgh. viii + 332 pages. SECOND EDITION. Crown 8vo, 10s. 6d. net.

PRACTICAL CHEMICAL PHYSIOLOGY. By W. W. TAYLOR, M.A., D.Sc. Crown 8vo. ORDINARY EDITION, 4s. 6d. net; INTERLEAVED EDITION, 5s. net.

LECTURES ON THEORETICAL AND PHYSICAL CHEMISTRY. By the late Dr. J. H. VAN'T HOFF. Translated by R. A. LEHFELDT, D.Sc. In three vols., demy 8vo, 29s. net, or separately: PART I. CHEMICAL DYNAMICS. 12s. net. PART II. CHEMICAL STATICS. 8s. 6d. net. PART III. RELATIONS BETWEEN PROPERTIES AND COMPOSITION. 8s. 6d. net.

ANALYTICAL MICROSCOPY. By T. E. WALLIS, B.Sc., Lecturer in Botany at the School of Pharmacy of the Pharmaceutical Society of Great Britain. viii + 150 pages. Illustrated. Crown 8vo, cloth, 6s. net.

AN INORGANIC CHEMISTRY. By H. G. DENHAM, M.A., D.Sc., Ph.D., Professor of Chemistry in the University of Cape Town. xii + 684 pages, with 144 diagrams. Crown 8vo, 12s. 6d. net.

INORGANIC CHEMISTRY. A Textbook for Schools. By E. J. HOLMYARD, B.A., Head of the Science Department, Clifton College. viii + 564 pages, with 119 diagrams and 10 plates. Crown 8vo, 6s. 6d.

A HANDBOOK OF ORGANIC ANALYSIS: QUALITATIVE AND QUANTITATIVE. By H. T. CLARKE, B.Sc., A.I.C., with an Introduction by Professor J. NORMAN COLLIE, F.R.S. xvi + 264 pages. THIRD EDITION. Crown 8vo, 6s. 6d. net.

THE CHEMISTRY OF BREADMAKING. By JAMES GRANT, M.Sc. Techn., F.I.C., F.C.S., Lecturer and Examiner in the Victoria University and School of Technology, Manchester; Head of the School of Bakery and Confectionery. THIRD EDITION. viii + 230 pages. Illustrated. Crown 8vo, 6s. net.

CONFECTIONERS' RAW MATERIALS. By JAMES GRANT, M.Sc. Techn., F.I.C., F.C.S. viii + 174 pages. Illustrated. Crown 8vo, 8s. 6d. net.

AN INTRODUCTION TO BACTERIOLOGICAL AND ENZYME CHEMISTRY. By GILBERT J. FOWLER, D.Sc., F.I.C., Professor of Bio-Chemistry in the Indian Institute of Science, Bangalore. viii + 328 pages. Illustrated. Crown 8vo, 8s. 6d. net.

INORGANIC CHEMISTRY FOR SCHOOLS. By W. M. HOOTON, M.A., M.Sc., Chief Chemistry Master at Repton School. 416 pages. Crown 8vo, 5s.

ELEMENTS OF INORGANIC CHEMISTRY. By the late W. A. SHENSTONE, F.R.S. SIXTH EDITION, revised and partly rewritten by R. G. DURRANT, M.A., Assistant Master, Marlborough College. Crown 8vo, 6s. 6d.

OUTLINES OF INORGANIC CHEMISTRY. With Special Reference to its Historical Development. By E. B. LUDLAM, D.Sc., F.C.S., late Head of Chemical Department, Clifton College. xvi + 368 pages. Illustrated. Crown 8vo, 5s.

MINING PHYSICS AND CHEMISTRY. By J. W. WHITAKER, B.Sc., Lecturer in the Mining Department, University College, Nottingham. xii + 268 pages. Crown 8vo, 9s. net.

QUALITATIVE ANALYSIS: IN THEORY AND PRACTICE. By P. W. ROBERTSON, M.A., Ph.D., Professor of Chemistry, Victoria University College, New Zealand, and D. H. BURLEIGH, M.Sc., D.I.C., Science Master, Christ's Hospital. Demy 8vo, 4s. 6d.

QUALITATIVE AND VOLUMETRIC ANALYSIS. By W. M. HOOTON, M.A., M.Sc. Demy 8vo, 3s. 6d. net.

EXERCISES IN CHEMICAL CALCULATION. By H. F. COWARD, D.Sc., and W. H. PERKINS, M.Sc. viii + 152 pages. Crown 8vo, [illegible]s. [illegible]d. net.

FIRST AID IN THE LABORATORY AND WORKSHOP. By A. A. ELDRIDGE, B.Sc., and H. V. A. BRISCOE, D.Sc. Cloth, 1s. 3d. net.

BIOLOGY

FOUNDERS OF OCEANOGRAPHY AND THEIR WORK. By Sir WILLIAM HERDMAN, C.B.E., D.Sc., F.R.S., Emeritus Professor in the University of Liverpool. xii + 340 pages, with 29 plates. Demy 8vo, cloth, 21s. net.

MANUAL OF ENTOMOLOGY. By H. MAXWELL LEFROY, M.A., Professor of Entomology in the Imperial College of Science and Technology. xvi + 552 pages. Fully Illustrated. Demy 8vo, cloth, 35s. net.

BRITISH HYMENOPTERA. By A. S. BUCKHURST, D.I.C., L. N. STANILAND, D.I.C., and G. B. WATSON, D.I.C., Imperial College of Science and Technology. With many illustrations. Crown 4to, 9s. net.

THE MECHANISM OF LIFE. In Relation to Modern Physical Theory. By J. JOHNSTONE, D.Sc., Professor of Oceanography in the University of Liverpool. xii + 248 pages, with 53 diagrams. Demy 8vo, 15s. net.

ANIMAL LIFE IN DESERTS. By P. A. BUXTON, M.A. xvi + 17[illegible] pages, with 14 plates. Demy 8vo, 10s. 6d. net.

AN INTRODUCTION TO THE STUDY OF THE PROTOZOA. With special reference to the Parasitic Forms. By the late E. A. MINCHIN, M.A., Ph.D., F.R.S. xii + 520 pages, 194 diagrams. Demy 8vo, 25s. net.

HUMAN EMBRYOLOGY AND MORPHOLOGY. By Sir ARTHUR KEITH, F.R.S., M.D., F.R.C.S., Conservator of the Museum and Hunterian Professor, Royal College of Surgeons, England. FOURTH EDITION. viii + 491 pages. Demy 8vo, 30s. net.

THE MIGRATIONS OF FISH. By ALEXANDER MEEK, M.Sc., F.L.S., F.Z.S., Professor of Zoology, Armstrong College, Newcastle-on-Tyne, and Director of the Dove Marine Laboratory. With illustrations and maps. xx + 428 pages. Demy 8vo, 18s. net.

ARBOREAL MAN. By F. WOOD JONES, M.B., D.Sc., Professor of Anatomy in the University of London. With 81 illustrations. Demy 8vo, 8s. 6d. net.

THE DEVELOPMENT OF BRITISH FORESTRY. By A. C. FORBES, F.H.A.S. xii + 274 pages, 70 illustrations. Demy 8vo, 10s. 6d. net.

THE EVOLUTION THEORY. By Dr. AUGUST WEISMANN, Professor of Zoology in the University of Freiburg in Breisgau. Translated by Professor J. ARTHUR THOMSON, and MARGARET THOMSON. Two Vols., xvi + 416 and viii + 396 pages, with over 130 illustrations. Royal 8vo, 32s. net.

MANUAL OF HUMAN PHYSIOLOGY. By LEONARD HILL, M.B., F.R.S. xii + 484 pages, with 177 illustrations. Crown 8vo, 6s.

ANIMAL BEHAVIOUR. By C. LLOYD MORGAN, LL.D., F.R.S., Professor of Psychology and Ethics in the University of Bristol. SECOND EDITION. viii + 344 pages, with 26 illustrations. Crown 8vo, 8s. 6d. net.

A HANDBOOK OF THE CONIFERÆ AND GINKGOACEÆ. By W. DALLIMORE and A. B. JACKSON. With 32 plates and 120 diagrams. Medium 8vo, cloth, 42s. net.

A CLASS-BOOK OF BOTANY. By G. P. MUDGE, A.R.C.Sc. (London), F.Z.S., Lecturer on Biology at the London Hospital Medical College, and A. J. MASLEN, F.L.S. xvi + 512 pages, with 228 illustrations. Crown 8vo, 8s. 6d. net.

封　底